Psychometrika

VOLUME XXII—1957

JANUARY–DECEMBER

PUBLISHED QUARTERLY

By THE PSYCHOMETRIC SOCIETY

AT 1407 SHERWOOD AVENUE
RICHMOND 5, VIRGINIA

Psychometrika

CONTENTS

VOLUME TWENTY-TWO MARCH 1957 NUMBER 1

SOUTHERN REGIONAL GRADUATE SUMMER SESSIONS IN STATISTICS

The fourth Southern Regional Graduate Summer Session in Statistics will be held from June 12 through July 20 at the Virginia Polytechnic Institute, Blacksburg, Virginia.

The session lasts six weeks and each course offered carries approximately five quarter hours of graduate credit. The summer work in statistics may be applied towards residence requirements at any one of the cooperating institutions, as well as certain other institutions, in partial fulfillment of residence requirements for graduate degrees.

The faculty for the 1957 session will include E. J. Williams, D. B. DeLury, J. L. McHugh, as well as the following staff members from the Virginia Polytechnic Institute: W. O. Ash, L. S. Brenna, R. A. Bradley, C. W. Clunies-Ross, J. E. Freund, R. J. Freund, B. Harshbarger, C. Y. Kramer, and R. L. Wine.

Of particular interest will be the lectures by D. B. DeLury on the Sampling of Biological Populations. Other courses to be offered include Analysis of Variance, Rank Order Statistics, Stochastic Processes, Probability, Statistical Inference, Theory of Least Squares, Statistical Methods, Engineering Statistics, and Sampling. Seminars, which will include many of the foremost statisticians in the eastern part of the United States, will be conducted each afternoon Monday through Thursday from 3:00 to 4:30. These seminars will be on some of the more recent research in the field of statistics.

The total tuition fee will be $38.00 for the six-week term. Doctoral courtesy will be offered to those holding doctoral degrees. Living and other expenses at the Virginia Polytechnic Institute are reasonable. The Virginia Polytechnic Institute is located on the scenic Alleghany Mountain plateau 2100 feet above sea level. The summer climate is delightful.

Inquiries shoud be addressed to Boyd Harshbarger, Head, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

NEW ANALYSIS OF VARIANCE FORMULAS FOR TREATING DATA FROM MUTUALLY PAIRED SUBJECTS*

JOSEPH LEV† AND ELAINE F. KINDER‡

YERKES LABORATORIES OF PRIMATE BIOLOGY, ORANGE PARK, FLORIDA

The experimental design considered in this paper is one in which each of a group of several subjects is observed in the presence of each of the other subjects of the group, the entire set of possible pairings being repeated or replicated on several occasions. Analysis of variance formulas are described for this somewhat unusual design. Both the model of constants and mixed model are considered. Reliability formulas growing out of the analysis of variance calculations are developed.

This paper describes new formulas in the analysis of variance. The novelty arises from the experimental design in which observations were made on pairs of chimpanzees of a group, each animal being paired with every other one. The formulas of analysis of variance usually given in textbooks are not appropriate in this situation because of the incompleteness of the layout. An animal cannot be paired with itself; consequently cells which correspond to this pairing are blank.

In the study for which this analysis of variance was developed, observations on all pairs were repeated five times at irregular intervals over a period of five years. Because of these repetitions it was possible to study interactions between pairs of animals as well as other interactions.

An analysis of variance for a design involving mutual pairings of a group of individuals when the observations are made once only is described by Quenouille [3, p. 256]. The analysis described by Quenouille does not consider repetitions or interactions. Although the analysis described in this paper was developed for an experiment dealing with chimpanzees, it is clearly applicable to experiments of the same design.

Each score in this study is a measure of the activity of a chimpanzee observed for a session of 20 minutes. During the session the chimpanzee was alone in a cage sufficiently large to permit free movement. The cage contained several items of equipment, such as two shelves, one high and one low, a strap suspended from the ceiling on which the animal could swing, and some toys. A second animal was placed in a small adjoining cage separated from the first cage only by a grating so that the two animals were in full view of each other.

*This work was supported in part by grants M-627; M-627C from the National Institutes of Health, Public Health Service.
†New York State Department of Civil Service.
‡New York State Department of Mental Hygiene.

A continuous second by second record was kept of the behavior movements of the first animal, whom we shall call the *protagonist*. No comparable record was kept of the behavior of the second animal, the *partner*, whose confining cage prevented escape from the grating, and limited its movement. Each protagonist was paired with six partners in successive observational sessions. The purpose of introducing the partner into the study was to ascertain whether the activity of the protagonist varied consistently with different partners.

Each of the seven chimpanzee subjects was paired with every other subject in the protagonist-partner relationship, making 42 pairings in each series of observations. The record for each session was later scored to obtain a summary activity measure, which was in essence a weighted average of scores for the various behavioral acts of the protagonist during the session.

Observations involving all 42 pairings were repeated several times over a period of five years, giving the five sets of observations which are analyzed in this paper. A set of sessions involving all 42 pairings and carried out at approximately the same time will be referred to as a *replication* of the experiment. The dates of the five replications were October 1943, November 1943, October 1944, April 1948 and July 1948.

The activity scores obtained in the first observational period, October 1943, are shown in Table 1. In this table the chimpanzees acting as protagonists are named in the column headings; their partners are named in the side headings of the rows. A score is the activity measure of the chimpanzee named in the heading of the column in which the score occurs when this chimpanzee is paired with the chimpanzee named in the corresponding row heading. For example, the score 252 in the column headed Falla is the activity measure of Falla in the presence of Banka. The activity measure of Banka in the presence of Falla is 189. The table immediately suggests analysis of variance as a natural method for analyzing the data. The complete analysis involves joint consideration of three factors and their interactions.

Factor 1: *Differences among protagonists.* This analysis should answer the question whether the observed differences among chimpanzees as protagonists are small enough to be attributed to chance or sufficiently great to be regarded as essential differences among the animals.

Factor 2: *Differences among partners.* This analysis is made to determine whether the observed differences in activity stimulated by partners are of such magnitude that they may be attributed to chance.

Factor 3: *Differences among replications.* Differences in activity among replications may be attributable to age of animals, to the time between the replications, or to their order.

The interactions among these factors are also of interest in the consideration of the data at hand.

Error. The chief error component used in the analysis of variance is the

second-order interaction. This is always the error component for testing significance of the interactions. However, the significance of the factors, or main effects, may also be tested by some of the first-order interactions.

The factors and their interactions will be shown in relation to the activity data of this study in several tables. A complete display of all scores would require five tables like Table 1, each showing scores for 42 pairings for a

TABLE 1

Activity Scores of Seven Chimpanzees Each

Paired with Every Other Chimpanzee

(October 1943)

	Protagonist							
Partner	Banka	Falla	Fanny	Flora	Jed	Jent	Karla	Total
Banka		252	199	168	197	165	148	1129
Falla	189		252	211	155	215	264	1286
Fanny	188	279		205	150	186	141	1149
Flora	160	259	206		195	155	144	1119
Jed	179	305	219	252		231	157	1343
Jent	178	266	255	211	150		203	1263
Karla	170	217	278	189	163	207		1224
Total	1064	1578	1409	1236	1010	1159	1057	8513

given replication, i.e., four tables in addition to Table 1. These four tables will be omitted. The entries in Tables 2, 3, and 4 are totals of entries based on the five basic tables.

In Table 2 each entry is the sum of the scores for five pairings of a particular protagonist with a particular partner, each of the five scores being a measure of activity at a given session. In Table 3 each entry is the sum of the activity scores of six protagonists paired with a given partner at a given replication. In Table 4 each entry is the sum of the activity scores of a protagonist paired with each of his six partners.

In order to visualize the relationship between the basic activity scores like those shown in Table 1 and the sums shown in Tables 2, 3 and 4, the

TABLE 2

Sums of Activity Scores for Five Pairings

of Protagonist and Partner

	Protagonist							
Partner	Banka	Falla	Fanny	Flora	Jed	Jent	Karla	Total
Banka		1021	996	757	787	776	898	5235
Falla	709		1123	825	644	756	981	5038
Fanny	799	1070		813	679	767	834	4962
Flora	714	1121	996		770	755	808	5164
Jed	800	1161	929	866		811	890	5457
Jent	808	1016	1118	848	694		917	5401
Karla	735	1123	1212	882	750	794		5496
Total	4565	6512	6374	4991	4324	4659	5328	36753

TABLE 3

Sums of Activity Scores of Six

Protagonists Paired with a Given Partner

Partner	Replication					Total
	Oct. 1943	Nov. 1943	Oct. 1944	April 1948	July 1948	
Banka	1129	950	1167	980	1009	5235
Falla	1286	911	993	977	871	5038
Fanny	1149	969	1126	826	892	4962
Flora	1119	1131	1081	934	899	5164
Jed	1343	1021	1065	1089	939	5457
Jent	1263	992	1232	995	919	5401
Karla	1224	1150	1173	1011	938	5496
Total	8513	7124	7837	6812	6467	36753

TABLE 4

Sums of Activity Scores for Each

Protagonist Paired with Six Partners

Replication	Protagonist							Total
	Banka	Falla	Fanny	Flora	Jed	Jent	Karla	
Oct. 1943	1064	1578	1409	1236	1010	1159	1057	8513
Nov. 1943	918	1143	1274	927	873	1043	946	7124
Oct. 1944	1060	1592	1143	1091	967	750	1234	7837
Apr. 1948	814	1053	1358	862	718	1018	989	6812
July 1948	709	1146	1190	875	756	689	1102	6467
Total	4565	6512	6374	4991	4324	4659	5328	36753

reader will note that the totals of the rows in Table 1 are recorded in the first column of Table 3, and the totals of the columns in Table 1 are the entries in the first row of Table 4. The remaining entries in Tables 2, 3, 4 are obtained similarly.

Formulas and computations for the analysis of variance will now be presented. The mathematical basis for the formulas will be sketched briefly in the mathematical appendix. The following symbolism will be needed for the formulas:

$X_{ijk} \equiv$ the basic activity score resulting from a single observation session in which the ith partner is paired with the jth protagonist in the kth replication.

For the study at hand i ranges from 1 through 7,

j ranges from 1 through 7,

k ranges from 1 through 5,

$i \neq j$ for the same observation.

$X_{ij.} \equiv$ the sum of all observations for the ith partner and the jth protagonist, the summation being made over all replications. This is an entry in Table 2.

$X_{i.k} \equiv$ the sum of all activity scores for the ith partner in the kth replication, the summation extending over all protagonists. This is an entry in Table 3.

$X_{.jk} \equiv$ the sum of all observations for the jth protagonist in the kth replication. This is an entry in Table 4.

$X_{i..} \equiv$ the sum of all observations on the ith partner for all protagonists and all replications. It is an entry in the column headed "Total" in Table 2.

$X_{.j.} \equiv$ the sum of all observations on the jth protagonist, the summation extending over all partners and all replications. It is an entry in the row headed "Total" in Table 2 (also Table 4).

$X_{..k} \equiv$ the sum of all observations in the kth replication. It is an entry in the row headed "Total" in Table 3.

$X_{...} \equiv$ the total of all observations in all replications.

$p \equiv$ the number of animals in the study.

$r \equiv$ the number of replications.

In the following formulas SS is an abbreviation for sum of squares, MS for mean square, df for degrees of freedom, and F is the usual ratio of mean squares. For variation among partners:

$$\text{(1)} \qquad \text{SS} = \frac{p-1}{rp(p-2)} \sum X_{i..}^2 + \frac{2 \sum X_{i..} X_{.i.}}{rp(p-2)} + \frac{\sum X_{.j.}^2}{rp(p-1)(p-2)} - \frac{X_{...}^2}{r(p-1)(p-2)}.$$

$$\text{df} = p - 1.$$

For variation among protagonists:

$$\text{(2)} \qquad \text{SS} = \frac{p-1}{rp(p-2)} \sum X_{.j.}^2 + \frac{2 \sum X_{i..} X_{.i.}}{rp(p-2)} + \frac{\sum X_{i..}^2}{rp(p-1)(p-2)} - \frac{X_{...}^2}{r(p-1)(p-2)}.$$

$$\text{df} = p - 1.$$

For variation among replications:

$$\text{(3)} \qquad \text{SS} = \frac{\sum X_{..k}^2}{p(p-1)} - \frac{X_{...}^2}{rp(p-1)}.$$

$$\text{df} = r - 1.$$

For partner × protagonist interaction:

$$\text{SS} = \frac{\sum\sum X_{ij.}^2}{r} - \frac{p-1}{rp(p-2)}\left(\sum X_{i..}^2 + \sum X_{.j.}^2\right) - \frac{2\sum X_{i..}X_{.i.}}{rp(p-2)} + \frac{X_{...}^2}{r(p-1)(p-2)}. \tag{4}$$

$$\text{df} = p^2 - 3p + 1.$$

For partner × replication interaction:

$$\text{SS} = \frac{p-1}{p(p-2)}\sum\sum X_{i.k}^2 + \frac{\sum\sum X_{.jk}^2}{p(p-1)(p-2)} + \frac{2\sum\sum X_{i.k}X_{.ik}}{p(p-2)} - \frac{(p-1)\sum X_{i..}^2}{rp(p-2)} - \frac{\sum X_{.j.}^2}{rp(p-1)(p-2)} - \frac{\sum X_{..k}^2}{(p-1)(p-2)} - \frac{2\sum X_{i..}X_{.i.}}{rp(p-2)} + \frac{X_{...}^2}{r(p-1)(p-2)}. \tag{5}$$

$$\text{df} = (p-1)(r-1).$$

For protagonist × replication interaction:

$$\text{SS} = \frac{p-1}{p(p-2)}\sum\sum X_{.jk}^2 + \frac{\sum\sum X_{i.k}^2}{p(p-1)(p-2)} + \frac{2\sum\sum X_{i.k}X_{.ik}}{p(p-2)} - \frac{(p-1)\sum X_{.j.}^2}{rp(p-2)} - \frac{\sum X_{i..}^2}{rp(p-1)(p-2)} - \frac{\sum X_{..k}^2}{(p-1)(p-2)} - \frac{2\sum X_{i..}X_{.i.}}{rp(p-2)} + \frac{X_{...}^2}{r(p-1)(p-2)}. \tag{6}$$

$$\text{df} = (p-1)(r-1).$$

For error:

$$\text{SS} = \sum\sum\sum X_{ijk}^2 - \frac{\sum\sum X_{ij.}^2}{r} - \frac{p-1}{p(p-2)}\sum\sum X_{i.k}^2 - \frac{p-1}{p(p-2)}\sum X_{.jk}^2 - \frac{2\sum\sum X_{i.k}X_{.ik}}{p(p-2)} + \frac{p-1}{rp(p-2)}\sum X_{i..}^2 + \frac{p-1}{rp(p-2)}\sum X_{.j.}^2 + \frac{\sum X_{..k}^2}{(p-1)(p-2)} + \frac{2\sum X_{i..}X_{.i.}}{rp(p-2)} - \frac{X_{...}^2}{r(p-1)(p-2)}. \tag{7}$$

$$\text{df} = (r-1)(p^2 - 3p + 1).$$

The formulas will be applied to computations relating to the activity data of the study. The sum of squares $\sum\sum\sum X_{ijk}^2$ is not available from the tables because the original scores are not shown. Additional sums of squares will be shown to help the reader, together with their source.

$$\begin{aligned}
\sum\sum\sum X_{ijk}^2 &= 6{,}830{,}715, \text{ from original measures,}\\
\sum\sum X_{ij.}^2 &= 33{,}060{,}491, \text{ from entries of Table 2,}\\
\sum\sum X_{i.k}^2 &= 39{,}164{,}035, \text{ from entries of Table 3,}\\
\sum\sum X_{.jk}^2 &= 40{,}314{,}363, \text{ from entries of Table 4,}\\
\sum\sum X_{i.k}X_{.ik} &= 38{,}728{,}822, \text{ from entries of Tables 3 and 4,}\\
\sum X_{i..}^2 &= 193{,}230{,}675, \text{ from row totals of Table 2,}\\
\sum X_{.j.}^2 &= 197{,}574{,}167, \text{ from column totals of Table 2,}\\
\sum X_{..k}^2 &= 272{,}866{,}547, \text{ from row totals of Table 4,}\\
\sum X_{i..}X_{.i.} &= 192{,}148{,}558, \text{ from totals of Table 2,}\\
X_{...}^2 &= 1{,}350{,}783{,}009.
\end{aligned}$$

The analysis of variance is shown in Table 5.

The analysis of variance in Table 5 is based on a model in which the components of a measure are constants, except for the error component. Consequently the conclusions are valid only for the seven animals in the study.

It is more interesting to study a second model in which the seven animals are regarded as a random sample from a population of similar animals. This is a mixed model described in the mathematical appendix. All components involving animals are then treated as random variables. Only the general mean and the component due to the main effect of replications are constants in this model. Using this model the interactions are still tested as in Table 5, but the main effects require more complex tests.

Tests for partner, or protagonist, main effects are made directly by the ratio of mean square of the main effect to the partner × protagonist interaction. Thus the test for significance of differences among protagonists is

$$F = 24{,}794.0/769.3 = 32.2,$$

which is still significant at the 0.5 per cent level.

The test for significance of variation among replications is more complex. Here

$$y_1 = \frac{(p-1)^2}{p(p-2)} \text{ (mean square of protagonist} \times \text{replication interaction).}$$

$$(8)\quad y_2 = \frac{(p-1)^2}{p(p-2)} \text{ (mean square of partner} \times \text{replication interaction).}$$

$$y_3 = \frac{p^2-2p+2}{p(p-2)} \text{ (mean square of error).}$$

Then

$$F = \frac{\text{mean square of replications}}{y_1 + y_2 - y_3}. \tag{9}$$

This ratio has an F distribution (approximately) with $r - 1$ degrees of freedom in the numerator and

$$n_2 = \frac{(y_1 + y_2 - y_3)^2}{\frac{y_1^2}{(r-1)(p-1)} + \frac{y_2^2}{(r-1)(p-1)} + \frac{y_3^2}{(r-1)(p^2-3p+1)}} \tag{10}$$

degrees of freedom in the denominator.

For the data in Table 5 the test for differences among replications indicates that $F = 5.25$, $n_1 = 4$, and $n_2 = 26$. This value of F also shows significance at the 0.5 per cent level.

The treatment of the replication component as a constant in the second model calls for justification. This treatment was adopted because the replications were spread over a long period, with the consequent changes which may be attributable to time. In fact the data show a fairly consistent decline in average activity over the replications. It would seem illogical, therefore, to treat the replications as a random sample from a population of replications. Even if the replications had been concentrated within a shorter period, one would hesitate to regard them as a random sample because of possible effects of order among the replications.

The second model leads to a measure of the reliability of the mean score of a protagonist over all his partners at a single replication, in our symbolism $X_{.ik}/(p - 1)$. The formula is developed in the appendix but will be described here.

Sample estimates of the variance components which arise from the second model are computed as follows. In the symbolism used, the subscripts a, b, c refer to partners, protagonists, and replications, respectively. A single subscript indicates a main effect or error. A pair of subscripts indicate an interaction. Using the mean squares derived from formulas (1) through (7),

$$s_e^2 = \text{MS error}, \tag{11}$$

$$s_{ac}^2 = \frac{(r-1)(p-1)}{rp(p-2)}(\text{MS partner} \times \text{replication} - s_e^2), \tag{12}$$

$$s_{bc}^2 = \frac{(r-1)(p-1)}{rp(p-2)}(\text{MS protagonist} \times \text{replication} - s_e^2), \tag{13}$$

$$s_{ab}^2 = \frac{1}{r}(\text{MS partner} \times \text{protagonist} - s_e^2), \tag{14}$$

$$s_b^2 = \frac{p-1}{rp(p-2)}(\text{MS protagonist} - \text{MS partner} \times \text{protagonist}), \tag{15}$$

$$(16) \qquad s_a^2 = \frac{p-1}{rp(p-2)} \text{(MS partner} - \text{MS partner} \times \text{protagonist)}.$$

The reliability of the mean of a protagonist over all partners at a single replication is then

$$(17) \quad \text{Replication reliability} = \frac{(p-1)s_b^2 + s_a^2 + s_{ab}^2}{(p-1)s_b^2 + s_a^2 + s_{ab}^2 + (p-1)s_{bc}^2 + s_{ac}^2 + s_e^2},$$

which is the correlation between mean scores for the same protagonists and partners at separate replications. From Table 5,

TABLE 5

Analysis of Variance of Activity Data

with Second-Order Interaction as Error

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Partners	3,580	6	596.7	1.05
Protagonists	148,764	6	24,794.0	43.57*
Replications	64,522	4	16,130.5	28.34*
Partner x Protagonist Interaction	22,311	29	769.3	1.35
Partner x Replication Interaction	19,349	24	806.2	1.42
Protagonist x Replication Interaction	66,289	24	2,762.0	4.85*
Error	66,012	116	569.1	

*significant at 0.5% level

$$s_e^2 = 569,$$

$$s_{ac}^2 = \frac{24}{175}(806 - 569) = 32,$$

$$s_{bc}^2 = \frac{24}{175}(2{,}762 - 569) = 301,$$

$$s_{ab}^2 = \frac{1}{5}(769 - 569) = 40,$$

$$s_b^2 = \frac{6}{175}(24{,}794 - 769) = 824,$$

$$s_a^2 = \frac{6}{175}(596 - 569) = 1,$$

$$\text{Replication reliability} = \frac{6(824) + 1 + 40}{6(824) + 1 + 40 + 6(301) + 32 + 569} = .674.$$

In addition to replication reliability, there is also partner reliability. This measure of reliability is based on the mean score of one protagonist with one partner over all replications, $X_{ii}./r$. The resulting partner relia-

bility is then

$$\text{(18)} \qquad \text{Partner reliability} = \frac{rs_b^2}{rs_b^2 + rs_a^2 + rs_{ab}^2 + s_e^2},$$

which is the correlation between the mean scores of protagonists computed over all replications but with differing partners. For the data in Table 5 this reliability is .842.

Where both replications and partners differ, but protagonists remain the same, the reliability is

$$\text{(19)} \qquad \text{Partner-replication reliability} = \frac{s_b^2}{s_b^2 + s_a^2 + s_{ab}^2 + s_{ac}^2 + s_{bc}^2 + s_e^2}.$$

For the data in Table 5 this reliability is .486.

Mathematical Appendix

It is the purpose of this appendix to state the mathematical basis for the formulas used in the paper.

The sums of squares were derived by use of the following model:

$X_{ijk} = m + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + e_{ijk}$, $\quad i \neq j$,
$X_{iik} = 0$
$e_{ijk} \equiv$ a normally distributed variable with mean 0 and variance σ^2 for all combinations of i, j, k. The remaining values m, a_i, b_j, etc. are constants satisfying the conditions stated below.
$m \equiv$ the mean of all observations,
$a_i \equiv$ the contribution of the ith partner,
$b_j \equiv$ the contribution of the jth protagonist,
$c_k \equiv$ the contribution of the kth replication,
$(ab)_{ij} \equiv$ the interaction of the ith partner with the jth protagonist,
$(ac)_{ik} \equiv$ the interaction of the ith partner with the kth replication,
$(bc)_{jk} \equiv$ the interaction of the jth protagonist with the kth replication.

The constants satisfy the following conditions:

$$\sum_{i=1}^{p} a_i = 0, \quad \sum_{j=1}^{p} b_j = 0, \quad \sum_{k=1}^{r} c_k = 0, \quad (ab)_{ii} = 0, \quad \sum_{i=1}^{p} (ab)_{ij} = 0$$

$$\sum_{j=1}^{p} (ab)_{ij} = 0, \quad \sum_{i=1}^{p} (ac)_{ik} = 0, \quad \sum_{k=1}^{r} (ac)_{ik} = 0, \quad \sum_{j=1}^{p} (bc)_{jk} = 0, \quad \sum_{k=1}^{r} (bc)_{jk} = 0$$

The determination of the sums of squares is carried out by least squares using methods described in several well-known references; see for example [1], [2] and [5]. A full derivation of formulas (1) through (7) is lengthy and will not be carried out here. However, a sketch of the method will be presented.

The likelihood ratio method described in sections 8.3 and 8.4 of [5], using particularly Theorem A, will be adopted. To apply this theorem set

up the sum of squares

$$U = \sum\sum\sum [X_{ijk} - m - a_i - b_j - c_k - (ab)_{ij} - (ac)_{ik} - (bc)_{jk}]^2.$$

Then find values of the constants which minimize U, subject to the conditions on the constants. If the constants found in this way by least squares are distinguished by circumflexes as $\hat{m}$, $\hat{a}_i$, etc., then

$$n\sigma_\Omega^2 = \sum\sum\sum [X_{ijk} - \hat{m} - \hat{a}_i - \hat{b}_j - \hat{c}_k - \widehat{(ab)}_{ij} - \widehat{(ac)}_{ik} - \widehat{(bc)}_{jk}]^2$$

is the error variance.

To test a hypothesis that certain of the constants are zero, similar sums of squares are formed with these constants omitted. For example, to test the hypothesis that the a_i are all zero, form the sum of squares

$$U_\omega = \sum\sum\sum [X_{ijk} - m - b_j - c_k - (ab)_{ij} - (ac)_{ik} - (bc)_{jk}]^2.$$

In terms of the least squares solutions, using the attendant conditions,

$$n\sigma_\omega^2 = \sum\sum\sum [X_{ijk} - \hat{\hat{m}} - \hat{\hat{b}}_j - \hat{\hat{c}}_k - \widehat{\widehat{(ab)}}_{ij} - \widehat{\widehat{(ac)}}_{ik} - \widehat{\widehat{(bc)}}_{jk}]^2.$$

Finally the sum of squares for partners is given by

$$n\sigma_\omega^2 - n\sigma_\Omega^2 .$$

The least squares estimates of the constants for substitution in $n\sigma_\Omega^2$ are

$$\hat{m} = \frac{X_{...}}{rp(p-1)},$$

$$\hat{a}_i = \frac{(p-1)X_{i..} + X_{.i.} - X_{...}}{rp(p-2)},$$

$$\hat{b}_j = \frac{(p-1)X_{.j.} + X_{j..} - X_{...}}{rp(p-2)},$$

$$\hat{c}_k = \frac{X_{..k}}{p(p-1)} - \frac{X_{...}}{rp(p-1)},$$

$$\widehat{(ab)}_{ij} = \frac{p(p-2)X_{ij.} - (p-1)(X_{i..} + X_{.j.}) - X_{.i.} - X_{j..} + \frac{pX_{...}}{p-1}}{rp(p-2)},$$

$$\widehat{(ac)}_{ik} = \frac{p-1}{p(p-2)} X_{i.k} + \frac{X_{.ik}}{p(p-2)} - \frac{p-1}{rp(p-2)} X_{i..} - \frac{X_{.i.}}{rp(p-2)} - \frac{X_{..k}}{p(p-2)} + \frac{X_{...}}{rp(p-2)},$$

$$\widehat{(bc)}_{jk} = \frac{p-1}{p(p-2)} X_{.jk} + \frac{X_{j.k}}{p(p-2)} - \frac{p-1}{rp(p-2)} X_{.j.} - \frac{X_{j..}}{rp(p-2)} - \frac{X_{..k}}{p(p-2)} + \frac{X_{...}}{rp(p-2)}.$$

The derivation of $n\sigma_\Omega^2$ is greatly simplified if one notes that it is equal to

$$n\sigma_\Omega^2 = \sum\sum\sum X_{ijk}[X_{ijk} - \hat{m} - \hat{a}_i - \hat{b}_j - \hat{c}_k - \widehat{(ab)}_{ij} - \widehat{(ac)}_{ik} - \widehat{(bc)}_{jk}].$$

The derivation of $n\sigma_\omega^2$, when the a_i are set equal to zero, leads to the following estimates of the constants:

$$\hat{\hat{m}} = \hat{m}, \quad \hat{\hat{c}}_k = \hat{c}_k, \quad \widehat{\widehat{(ab)}}_{ij} = \widehat{(ab)}_{ij},$$

$$\widehat{\widehat{(ac)}}_{ik} = \widehat{(ac)}_{ik}, \quad \widehat{\widehat{(bc)}}_{jk} = \widehat{(bc)}_{jk},$$

but

$$\hat{\hat{b}}_j = \frac{X_{.j.}}{r(p-1)} - \frac{X_{...}}{rp(p-1)}.$$

In deriving these estimates, the Lagrange multipliers, by which the conditions on the constants are introduced, play a very important part, and the mathematics becomes somewhat tricky. For example, obtaining the estimate $\widehat{\widehat{(ab)}}_{ij}$ involves the expression

$$\sum\sum\sum [X_{ijk} - m - b_j - c_k - (ab)_{ij} - (ac)_{ik} - (bc)_{jk}]^2 - 2\sum_i \lambda_{i..} \sum_j (ab)_{ij} - 2\sum_j \lambda_{.j.} \sum_i (ab)_{ij}.$$

Noting that $m + b_j = X_{.j.}/r(p-1)$, and using the conditions on the remaining constants, then the least squares equations for evaluating the $(ab)_{ij}$ are

$$r(ab)_{ij} + \lambda_{i..} + \lambda_{.j.} = X_{ij.} - \frac{X_{.j.}}{p-1}.$$

The Lagrange multipliers are not zero here as they are in many of the other derivations, but they can be eliminated by considering the related equations

$$r(ab)_{ji} + \lambda_{j..} + \lambda_{.i.} = X_{ji.} - \frac{X_{.i.}}{p-1}.$$

Elimination of the Lagrange multipliers and evaluation of the $(ab)_{ij}$ is accomplished by use of the following relations:

$$\sum_{i \neq j} \lambda_{i..} + (p-1)\lambda_{.j.} = 0,$$

$$(p-1)\lambda_{i..} + \sum_{j \neq i} \lambda_{.j.} = \frac{(p-1)X_{i..} + X_{.i.} - X_{...}}{p-1},$$

$$\sum_{j \neq i} \lambda_{j..} + (p-1)\lambda_{.i.} = 0,$$

$$(p-1)\lambda_{j..} + \sum_{i \neq j} \lambda_{.i.} = \frac{(p-1)X_{j..} + X_{.j.} - X_{...}}{p-1},$$

and substitution of these relations, and the preceding ones, in the equations:

$$\sum_{i=1}^{p} \lambda_{i..} + \sum_{j=1}^{p} \lambda_{.j.} = \sum_{i \neq j} \lambda_{i..} + \lambda_{j..} + \sum_{j \neq i} \lambda_{.j.} + \lambda_{.i.} = 0,$$

and in the parallel equation:

$$\sum_{j \neq i} \lambda_{j..} + \lambda_{i..} + \sum_{i \neq j} \lambda_{.i.} + \lambda_{.j.} = 0.$$

Considerations similar to the above lead to the remaining sums of squares.

The reader should note that the sums of squares in formulas (1) through (7) do not total to the sum of squares of deviations from the grand mean of all observations. This is a characteristic of the nonorthogonal case in analysis of variance, of which the present analysis is an example.

In order that the seven animals may be regarded as a random sample from a population of similar chimpanzees, the model is extended so that all components involving the animals will be variables which are normally and independently distributed with mean zero and variances to be defined. Using the notation $V(X)$ for variance of X, we shall write

$$V(a_i) = \sigma_a^2, \qquad V(b_j) = \sigma_b^2, \qquad V[(ab)_{ij}] = \sigma_{ab}^2$$
$$V[(ac)_{ik}] = \sigma_{ac}^2, \qquad V[(bc)_{jk}] = \sigma_{bc}^2, \qquad V(e_{ijk}) = \sigma_e^2;$$

m and c_k are constants as before.

The following conditions are imposed on the components:

$$\sum_k c_k = 0, \qquad \sum_k (ac)_{ik} = 0 \quad \text{for each } i,$$

$$\sum_k (bc)_{jk} = 0 \quad \text{for each } j.$$

The mean squares using this model are derived from formulas (1) through (7). The expected values of the mean squares are

$$E(\text{MS partners}) = \sigma_e^2 + r\sigma_{ab}^2 + \frac{rp(p-2)}{p-1}\sigma_a^2$$

$$E(\text{MS protagonists}) = \sigma_e^2 + r\sigma_{ab}^2 + \frac{rp(p-2)}{p-1}\sigma_b^2$$

$$E(\text{MS replications}) = \sigma_e^2 + \frac{r(p-1)}{r-1}(\sigma_{ac}^2 + \sigma_{bc}^2) + \frac{\sum c_k^2}{r-1}$$

$$E(\text{MS partner} \times \text{protagonist}) = \sigma_e^2 + r\sigma_{ab}^2$$

$$E(\text{MS partner} \times \text{replication}) = \sigma_e^2 + \frac{rp(p-2)}{(r-1)(p-1)}\sigma_{ac}^2$$

$$E(\text{MS protagonist} \times \text{replication}) = \sigma_e^2 + \frac{rp(p-2)}{(r-1)(p-1)}\sigma_{bc}^2$$

$$E(\text{MS error}) = \sigma_e^2.$$

The formulas for testing main effects against interactions, (8) through (10), are derived from these expectations by methods described in [1, chap. 28]. From these expectations the reliability formulas shown above can also be derived. The values of s_e^2 , s_{ac}^2 , s_{bc}^2 , etc., are obtained by solving the expressions for the expectations for the variances σ_e^2 , σ_{ac}^2 , σ_{bc}^2 , etc., successively, and then substituting sample values for the population values.

To derive the formula for replication reliability, from the second model

$$\frac{X_{.jk}}{p-1} = m + \frac{\sum_i a_i}{p-1} + b_j + c_k + \frac{\sum_i (ab)_{ij}}{p-1} + \frac{\sum_i (ac)_{ik}}{p-1} + (bc)_{jk} + \frac{\sum_i e_{ijk}}{p-1}.$$

Consider now the mean for a protagonist at a parallel replication k' so that

$$\frac{X_{.jk'}}{p-1} = m + \frac{\sum_i a_i}{p-1} + b_j + c_{k'} + \frac{\sum_i (ab)_{ij}}{p-1} + \frac{\sum_i (ac)_{ik'}}{p-1} + (bc)_{jk'} + \frac{\sum_i e_{ijk'}}{p-1}.$$

By the model

$$E\left(\frac{X_{.jk}}{p-1}\right) = m + c_k , \qquad E\left(\frac{X_{.jk'}}{p-1}\right) = m + c_{k'} .$$

Using these expectations and noting that $X_{.jk}/(p-1)$ and $X_{.jk'}/(p-1)$ differ only in the terms containing k and k' as subscripts,

$$\operatorname{cov}\left(\frac{X_{.jk}}{p-1}, \frac{X_{.jk'}}{p-1}\right) = \sigma_b^2 + \frac{\sigma_a^2}{p-1} + \frac{\sigma_{ab}^2}{p-1}.$$

By the model the expectation of a square of a variable is its variance, whereas the expectation of a product of two different variables is zero. Also

$$V\left(\frac{X_{.jk}}{p-1}\right) = V\left(\frac{X_{.jk'}}{p-1}\right) = \sigma_b^2 + \frac{\sigma_a^2}{p-1} + \frac{\sigma_{ab}^2}{p-1} + \sigma_{bc}^2 + \frac{\sigma_{ac}^2}{p-1} + \frac{\sigma_e^2}{p-1}.$$

The covariance and the variances jointly lead to the replication reliability (17).

To obtain the formula for partner reliability, by the second model

$$\frac{X_{ij.}}{r} = m + a_i + b_j + (ab)_{ij} + \frac{\sum_k e_{ijk}}{r} ,$$

since in summing over all replications the terms containing k vanish.

Similarly, for a different partner

$$\frac{X_{i'j\cdot}}{r} = m + a_{i'} + b_j + (ab)_{i'j} + \frac{\sum e_{ijk}}{r}.$$

Formula (18) follows readily.

To obtain formula (19) correlate

$$X_{ijk} = m + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + e_{ijk}$$

with the corresponding expression for $X_{i'jk'}$, where i and k differ from i' and k'.

REFERENCES

[1] Kempthorne, O. The design and analysis of experiments. New York: Wiley, 1952.
[2] Kolodzieczyk, S. On an important class of statistical hypotheses. *Biometrika*, 1935, **27**, 161-190.
[3] Quenouille, H. M. Design and analysis of experiments. New York: Hafner, 1952.
[4] Walker, H. M. and Lev, J. Statistical inference. New York: Holt, 1953.
[5] Wilks, S. S. Mathematical statistics. Princeton: Princeton Univ. Press, 1944.

Manuscript received 10/11/55.

Revised manuscript received 2/10/56.

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

ON THE RANKING PROBLEM

ISADORE BLUMEN

CORNELL UNIVERSITY

Observed rankings of objects can be treated as arising from a time dependent probability process. Under such circumstances, associations observed are an indication of the character of this underlying process. In the particular example treated in some detail here, a quantity related to Kendall's tau is found to have an important role and its properties are examined.

1. *Introduction*

One of the frequent problems which research workers in the social sciences raise with statisticians involves relations between variates when the observations consist of rankings of the units which are observed. There are many procedures for testing the hypothesis that the rankings are independent in a probability sense. But the research worker has usually anticipated this possibility—he is really interested in measuring the concordance of these rankings. On this point the measures hitherto proposed have not proved completely satisfactory. Speaking generally, they suffer in varying degrees from difficulties of interpretation, from stringent assumptions about underlying measurements, or from the lack of an adequate sampling theory.

It is the purpose of this paper to suggest the usefulness of probability processes as the point of departure for some problems of this sort. Instead of viewing a particular ranking as one out of a hypothetically infinite set of random drawings of objects from some population, suppose that there is some process producing the possible rankings, and that each ranking has its probability determined by the character of the process. The observed rankings of the n objects then give some insight into the process which is producing the concordance.

A particularly simple model will be used. Those who find it unrealistic and unreasonable as an approximation to their problems may try more complex ones. They are warned, however, that computational difficulties can easily become overwhelming with no corresponding gain in usefulness.

2. *Single Judge and a Standard Ranking: The Model*

There are a number of ways to formulate the problem. For convenience, it will be assumed that a judge is asked to rank a collection of n objects which have some natural order. These objects could be students who are

being ranked on ability, houses which are to be ranked on liveability, automobiles on beauty, and so on. The judge will generally begin with a tentative ranking and then compare objects whose rankings are close. He will adjust his rankings by comparison of close objects until satisfied or until required to report his ranking.

In this model, it is assumed that at each comparison the judge takes only those $(n - 1)$ pairs which differ by one rank and that he chooses one among these pairs at random, i.e., with probability $(n - 1)^{-1}$ for each pair. It is assumed that there is a preferred or *standard* order for the pair, and that the pair is put into this order with probability p, $0 \leq p \leq 1$, and in the reverse order with probability $q = 1 - p$. This process is repeated over and over—selecting one among the $(n - 1)$ adjacent pairs at random and assigning them the standard order with probability p without regard to the previous ordering.

This is a Markov process. The essential probability characteristics which determine the shift from one ranking to another can be given by a transition matrix, each element in the transition matrix being the probability that a ranking in one order will be changed to another order at each stage of the process.

Before writing this transition matrix, it will be convenient to group the $n!$ orders into classes. The first class, S_0 , will have one ranking, the objects in a standard order. The second class, S_1 , will consist of the $n - 1$ possible rankings obtained by permuting adjacent objects in the ranking S_0 . If adjacent objects are then permuted in rankings in the class S_1 , either the member of S_0 or a new ordering is obtained. All such new orderings form the class S_2 . Similarly S_3 is formed by permuting adjacent objects in S_2 and taking those rankings not in S_1 ; S_4 is formed from S_3 , and so on. It can be shown that the number of such classes is $[n(n - 1)/2] + 1$.

The transition matrix will be an $n!$ by $n!$ matrix in which the element in the ith row and jth column will be the probability that the ith of the $n!$ possible rankings of n objects will be changed by the judge into the jth of the possible rankings. The matrix will be written so that the first row will give the probability that a judge, given the objects in the standard order (in class S_0), will change to any other order or continue to use the ranking S_0 . The first column, on the other hand, will give the probability that he will move from any ranking into S_0 .

The second through the nth rows will give the probability of moving to any ranking from any one of the rankings in S_1 . The subscripts $i = 2, \cdots, n$ may be assigned in any arbitrary way to identify the rankings in S_1 ; corresponding subscripts are used for columns 2 through n in order to give the probabilities of moving into the corresponding rankings in S_1 . The next rows belong to rankings in the class S_2 , and so on to the one inverse ranking in $S_{n(n-1)/2}$.

The probability of moving from S_0 into any particular ranking in S_1 is $q(n-1)^{-1}$. There is no possibility of moving into any other ranking in a single comparison, and all other transition probabilities in the first row, except the one in the first column, designated m_{11}, are zero. Since the sum of the elements in any row must be unity, $m_{11} = p$.

In comparing pairs when the ranking is in S_1, there are three possibilities: (i) If the one pair which is not in the natural order is reversed, then the new ranking will be in S_0. The probability of this is $p(n-1)^{-1}$. (ii) If any other pair is reversed then the new ranking will be in S_2. The probability of moving from one of the rankings in S_1 to any particular ranking in S_2 which can be reached from it by one permutation is $q(n-1)^{-1}$. (iii) There will be no change with probability $[q + (n-2)p](n-1)^{-1}$, since the row sum must total to unity and all other probabilities in the second through the nth rows must be zero.

In general, the $n!$ by $n!$ probability matrix is partitioned into $\{[n(n-1)/2] + 1\}^2$ sub-matrices corresponding to the classes described above. All sub-matrices off the principal diagonal of sub-matrices by more than one class have zero elements since one permutation moves a ranking no more than one class at a time.

From any ranking, call it the ith, there are $(n-1)$ possible other rankings which can be reached by one permutation. There will be some number c_i $(0 \leq c_i \leq n-1)$ of these which lead to the class with a subscript which is one lower than that of the ith ranking. The probability of arriving at any particular one of these c_i rankings is $p(n-1)^{-1}$.

There will be $n - c_i - 1$ possible rankings in the class with the next higher subscript than the ith ranking. The probability of any particular one of these is $q(n-1)^{-1}$. Finally, the probability that there will be no change in a single comparison is

$$m_{ii} = [c_i q + (n - c_i - 1)p](n-1)^{-1}. \tag{2.1}$$

All other elements in the row corresponding to this ranking are zero.

For example, for $n = 3$,

$$M(p) = \left[\begin{array}{c|cc|cc|c} p & q/2 & q/2 & 0 & 0 & 0 \\ \hline p/2 & (q+p)/2 & 0 & q/2 & 0 & 0 \\ p/2 & 0 & (q+p)/2 & 0 & q/2 & 0 \\ \hline 0 & p/2 & 0 & (q+p)/2 & 0 & q/2 \\ 0 & 0 & p/2 & 0 & (q+p)/2 & q/2 \\ \hline 0 & 0 & 0 & p/2 & p/2 & q \end{array}\right], \tag{2.2}$$

where $M(p)$ is the transition matrix whose probabilities are determined by p.

For $p = 1/2$ and all n, $M(p)$ is symmetric. Furthermore, it is always possible after finitely many stages of the process—i.e., after finitely many comparisons—for any of the possible rankings to arise. Hence, from the theory of Markov chains [1] if a sufficiently large number of comparisons is to be made the probability of having any ranking will be the same as that for having any other ranking, i.e., $(n!)^{-1}$. More precisely, when $p = 1/2$ the limit of the probability that a judge will have the objects ranked in a particular way is $(n!)^{-1}$ no matter what the initial tentative ranking.

When $p \neq 1/2$, the limit probabilities can be obtained from the theory of Markov chains by solving for t_i in

$$\sum_{i=1}^{n!} t_i m_{ij} = t_j , \qquad (j = 1, 2, \cdots, n!) \tag{2.3}$$

and

$$\sum_{i=1}^{n!} t_i = 1,$$

where m_{ij} is the element in the ith row and jth column of $M(p)$. (When p is zero or one there is a trivial special case which is excluded hereafter.)

In particular, for $j = 1$,

$$pt_1 + t_2 p(n-1)^{-1} + \cdots + t_n p(n-1)^{-1} = t_1 . \tag{2.4}$$

Rewriting,

$$qt_1 = (n-1)^{-1} p(t_2 + t_3 + \cdots + t_n), \tag{2.5}$$

for which the relationship $t_{k_1} = qp^{-1}t_1$ is a solution when $k_1 = 2, 3, \cdots, n$.

For $j = 2, 3, \cdots, (n! - 1)$, the equations (2.3) can be typified by the one for $j = 2$. This equation is

$$\begin{aligned} t_1 q(n-1)^{-1} + t_2[q + (n-2)p](n-1)^{-1} \\ + t_{n+1} p(n-1)^{-1} + \cdots + t_{2n-2} p(n-1)^{-1} = t_2 . \end{aligned} \tag{2.6}$$

By straightforward substitution in (2.6), it is seen that $t_1 = pq^{-1}t_2$ and $t_{k_2} = qp^{-1}t_2$ are solutions, where $k_2 = n + 1, n + 2, \cdots, 2n - 2$.

There are, in general, more than $(n - 2)$ rankings in S_2 . It is clear, however, that solutions of (2.3) for any ranking in S_2 are $t_{k_s} = q^2 p^{-2} t_1$, where k_2 ranges over the rankings in S_2 . Similarly, for any ranking in S_3 $t_{k_s} = q^3 p^{-3} t_1$ are solutions to (2.3) and generally, $t_{k_s} = q^s p^{-s} t_1$, where k_s ranges over the states in the class with subscript s. If the condition that $\sum_{i=1}^{n!} t_i = 1$ is added, the above solutions are unique.

The limit probabilities which are the solution to (2.3) determine the behavior of the observed rankings. It is this probability distribution (see Tables 1 and 2 below for $n = 3$ and 4) which is required for the empirical

study of the behavior of the process described previously. Things are simplified further if it is noted that, for fixed s and p, the limit probabilities of all members in the class S_s, $s = 0, 1, 2, \cdots, n(n-1)/2$, are the same. That is, s is a sufficient statistic for p, and we can restrict ourselves to the distribution of s.

When $p = 1/2$, this distribution is basically that of Kendall's tau, since $\tau = 1 - 4sn^{-1}(n-1)^{-1}$. The probability generating function for this case is essentially given by Kendall [2] and is

$$\begin{aligned} E(x^s) &= (n!)^{-1}(1+x)(1+x+x^2)\cdots(1+x+\cdots+x^{n-1}) \\ &= [(n!)(1-x)^n]^{-1}\prod_{i=1}^{n}(1-x^i), \quad |x|<1. \end{aligned} \tag{2.7}$$

Hence, from (2.7) and the solutions to the equations like (2.6), the moment generating function of s is, for $0 < p < 1$,

$$\begin{aligned} \phi(\theta) &= E(e^{\theta s}) \\ &= \prod_{i=1}^{n-1}\frac{1+re^{\theta}+r^2e^{2\theta}+\cdots+r^ie^{i\theta}}{1+r+r^2+\cdots+r^i} \\ &= \left(\frac{1-r}{1-re^{\theta}}\right)^n\prod_{i=1}^{n}\frac{1-(re^{\theta})^i}{1-r^i}, \end{aligned} \tag{2.8}$$

where $|re^{\theta}| < 1$, and $r = qp^{-1}$.

Before proceeding further, it should be noted that the process discussed here is not the only process which gives limit distributions whose generating functions are (2.7) and (2.8). For example, any Markov process whose transition matrix had column totals equal to unity would give (2.7). Similarly, other simple processes lead to (2.8). (See for example [3].) Such processes have not been explored here because none seemed more reasonable simple approximations to useful situations than the one treated. And finally, as pointed out earlier, more complex distributions can also follow from other assumptions, but the exploration of these was beyond the purposes of this paper.

3. *Single Judge and Standard Ranking: Inference Problems*

Tests which involve τ in the usual way have optimum properties. Thus, to test the null hypothesis that an observed ranking arose from a process in which there is no tendency toward conformity with a natural ranking, i.e., $p = 1/2$, against alternatives that there is a tendency toward conformity, $p > 1/2$, a uniformly most powerful test is to reject the null hypothesis whenever τ is small. Similar remarks apply to invariant tests of $p = 1/2$ against $p \neq 1/2$.

The more interesting problems involve the estimation of the error, i.e.,

the estimation of p. It will first be shown that τ is not a satisfactory estimator of error when the process above is relevant.

Using the moment generating function of s (2.8),

$$\phi_0'(\theta) = \phi_0 \frac{\partial \log \phi}{\partial \theta}\bigg|_{\theta=0} = \frac{nr}{1-r} - \sum_{j=1}^{n} \frac{jr^j}{1-r^j}. \tag{3.1}$$

(We restrict ourselves to $r < 1$. The case $r > 1$ goes through in a parallel way, and the case $r = 1$ is treated by taking advantage of the continuity of the moment generating function.) Now,

$$E(\tau) = 1 - \frac{4}{n(n-1)} E(s). \tag{3.2}$$

(3.2) can be bounded by

$$E(\tau) - 1 < \frac{-4r}{(n-1)(1-r)} + \frac{4r}{n(n-1)(1-r)^3} \tag{3.3}$$

and

$$E(\tau) - 1 > \frac{-4r}{(n-1)(1-r)}. \tag{3.4}$$

Hence, $\lim_{n=\infty} E(\tau) = 1$ and, since τ must be less than unity, $\operatorname{plim}_{n=\infty} \tau = 1$ whenever $r < 1$. Similarly, it can be shown that when $r > 1$, $\operatorname{plim}_{n=\infty} \tau = -1$. When $r = 1$, we can obtain $E(\tau) = 0$ and compute the variance by using (2.8) and L'Hospital's rule. It follows that $\operatorname{plim}_{n=\infty} \tau = 0$ whenever $r = 1$.

Why τ has the peculiar property of converging to only these three values is seen from (2.7) and (2.8). For $r < 1$, the distribution of s is that of a convolution of n independent truncated geometric distributions, each with parameter r. Hence, s/n will have the ordinary sort of behavior and, in particular, will have as its limit the normal distribution whose mean is

$$E(s/n) = (1-p)(2p-1)^{-1}, \tag{3.5}$$

and whose variance is

$$\operatorname{var}(s/n) = p(1-p)(2p-1)^{-2}n^{-1}. \tag{3.6}$$

When $r = 1$, the distribution of s is that of the convolution of n independent chance variables, each having a point rectangular distribution, i.e., each chance variable x_i has the distribution $P\{x_i = j\} = (1+i)^{-1}$ for $j = 0, 1, 2, \cdots, i$. Hence, $\sum_{i=1}^{n} x_i = s$ must be divided by a term of order n^2, as is done in computing τ, if the mean of the limiting distribution is to be a bounded positive quantity.

Since τ is not a satifactory measure, attention is turned to estimation procedures for p and functions of p. The problem will be complicated by

the difficulty in writing the probability distribution for s explicitly. For small n, however, Table 16.3 in Kendall [1] can be used to write the exact probabilities, comparing various estimation procedures in this way, while for large n the asymptotic normality of s can be used.

A reasonable requirement for an estimate of p when n is small might well be unbiasedness. The method of obtaining unbiased estimates can be illustrated for $n = 3$. The probability that $s = 0$, when $r < 1$, is

$$\begin{aligned} P\{s = 0\} &= \frac{(1 - r)^2}{(1 - r^2)(1 - r^3)} \\ &= t_0 , \\ P\{s = 1\} &= 2rt_0 , \\ P\{s = 2\} &= 2r^2 t_0 , \\ P\{s = 3\} &= r^3 t_0 . \end{aligned}$$

Hence, if p_0^* is the estimator used when $s = 0$, p_1^* when $s = 1$, and so on, then from the definition of unbiasedness

$$p_0^* t_0 + 2p_1^* r t_0 + 2p_2^* r^2 t_0 + p_3^* r^3 t_0 = p = (1 + r)^{-1}.$$

Equating coefficients of the like powers of r,

$$p_0^* = 1, \qquad p_1^* = 1/2, \qquad p_2^* = 1/2, \qquad p_3^* = 0.$$

It is not difficult to show that these estimates are unbiased for all values of p.

In general, let a_s be the coefficient of x^s in the polynomial $(1 + x)(1 + x + x^2) \cdots (1 + x + \cdots + x^{n-1})$ and b_s be the coefficient of x^s in $(1 + x + x^2)(1 + x + x^2 + x^3) \cdots (1 + x + \cdots + x^{n-1})$. Then the unbiased estimates are

$$\begin{aligned} p_s^* &= \frac{b_s}{a_s} \\ &= \frac{b_s}{b_{s-1} + b_s} , \qquad \text{writing } b_{-1} = 0. \end{aligned} \tag{3.7}$$

This estimation procedure has the property of consistency but as n becomes large the difficulties in obtaining b_s are the same as those for a_s. A somewhat easier consistent estimate might, therefore, be used. One such estimate is

$$\begin{aligned} &\text{(a)} \quad \text{when } \frac{s}{n(n-1)} < 1/4 \text{ and } n > 2, \quad \hat{p}_{(a)} = \frac{n + s - 3}{n + 2s - 3}; \\ &\text{(b)} \quad \text{when } \frac{s}{n(n-1)} > 1/4 \text{ and } n > 2, \quad \hat{p}_{(b)} = 1 - \frac{n + s^* - 3}{n + 2s^* - 3}; \\ &\text{(c)} \quad \text{when } \frac{s}{n(n-1)} = 1/4 \text{ or } n = 2, \quad \hat{p}_{(c)} = 1/2; \end{aligned} \tag{3.8}$$

where

$$s^* = \frac{n(n-1)}{2} - s.$$

This estimation procedure is based on the fact that s has, approximately, a negative binomial distribution and (a) and (b) are essentially maximum likelihood estimates when $r < 1$ and $r > 1$, respectively, with a correction to reduce bias for small n. Since, when n is large, the conditions for use of (a) and (b) hold with probability arbitrarily close to unity the estimates are consistent for $r \neq 1$.

For $r = 1$, (a) and (b) will be used with equal probability. If (a) is used, then the conditional probability

$$P\left\{|\hat{p} - p| < \epsilon \middle| \frac{s}{n(n-1)} < \frac{1}{4}\right\} \to 0 \tag{3.9}$$

since

$$E(s) = \frac{n(n-1)}{4} \quad \text{and} \quad \text{var}\,(s) = \frac{n(n-1)(2n+5)}{72}.$$

A similar relation holds for (b), and together with (c), proves consistency.

The efficiency of these estimates is unknown. The probability distributions of s for $n = 3$ and $n = 4$ are given below for three values of p together with the estimates p^* and their variance. (Our estimate $\hat{p}$ coincides with p^* for these values of n.)

When $p \neq 1/2$ and n is large, the estimator is approximately normally distributed and it can be shown that

$$\text{var}\,(\hat{p}) \sim p(1-p)(2p-1)^2/n. \tag{3.10}$$

TABLE 1

Probability Distribution of s
and Estimate of p for n = 3

s	$p^* = \hat{p}$	$P\{s\} = P\{\hat{p}\}$		
		p = 1/2	p = 2/3	p = 3/4
0	1.0	.167	.381	.519
1	0.5	.333	.381	.346
2	0.5	.333	.190	.115
3	0.0	.167	.048	.019
Var (p*)		.084	.075	.072

TABLE 2

Probability Distribution of s
and Estimate of p for n = 4

s	$p^* = \hat{p}$	P{s} = P{$\hat{p}$}		
		p = 1/2	p = 2/3	p = 3/4
0	1.00	.042	.203	.350
1	0.67	.125	.305	.350
2	0.60	.208	.254	.195
3	0.50	.250	.152	.078
4	0.40	.208	.064	.022
5	0.33	.125	.019	.004
6	0.00	.042	.003	.000
Var (p*)		.032	.033	.038

This approximation is useful only when n is at least moderately large. Bias seems to be relatively unimportant. For $n = 6$ and $p = 2/3$, for example, there is an upward bias of .016.

When $p = 1/2$ there is no bias. In addition, it can be shown that when n is large p is approximately $1/2 \pm 1/n$.

4. *Several Processes*

Now consider the case where there are several processes, each with the same standard ranking. In practice, these processes may arise from m judges attempting to rank n students on intelligence, n cities which have m different indices of their development, and so on. The standard ranking is unknown to the statistician. He is asked to estimate this ranking and to give some indication of the usefulness of the judges or indices for ranking purposes.

Each of the m estimates of the standard ranking is given independently and assumed to be derived by a process which gives the probability distribution of section 2. The probability of any particular set of m rankings is, therefore,

$$(4.1) \qquad P\{s_{1k_1}\}P\{s_{2k_2}\} \cdots P\{s_{mk_m}\} = \prod_{i=1}^{m} \prod_{j=1}^{n} \frac{(1 - r_i)}{(1 - r_i^j)} r_i^{s_{ik_i}}, \qquad (r_i < 1)$$

where $P\{s_{ik_i}\}$ is the probability that the ranking arising from process i ($i = 1, 2, \cdots, m$), for which r_i is the error parameter of the ith judge, is in the class which is s_{ik_i} permutations from the standard ranking.

In particular, if $r_1 = r_2 = \cdots = r_n = r$,

$$\prod_{i=1}^{m} P\{s_{ik_i}\} = \prod_{j=1}^{n} \frac{(1-r)^m}{(1-r^j)^m} r^s, \tag{4.2}$$

where $s = s_{1k_1} + s_{2k_2} + \cdots + s_{mk_m}$ with the restriction that $r < 1$.

A maximum likelihood estimate of the standard ranking is obtained if a ranking which minimizes the total number of permutations from the estimated standard order is chosen. This is very simple in practice. For each process, assign rank 1 to the "best" object, rank 2 to the next, and so on to rank n for the last. Obtain the total (or average) score for each object over all processes. A maximum likelihood estimate of the natural order is given by assigning 1st rank to the object with the lowest score, 2nd to the object with next lowest score, and so on to the nth rank for the object with the highest total score. In the event that there are ties, this application of the method of maximum likelihood does not give a unique answer, and a randomization procedure which gives equal probability among the possible rankings of tied objects seems reasonable.

It is clear that, for fixed n, the standard ranking will tend to be chosen with probability one as m increases. For small m, however, the probability that a particular ranking will be chosen as the standard ranking does not seem to be easily obtainable in general.

For example, for $n = 3$ and $m = 2$, the probability of a correct ranking is $[(1 + r)(1 + r + r^2)]^{-1}$, the probability of either of the rankings one permutation from the standard order is $r(1 + 2r)[(1 + r)(1 + r + r^2)^2]^{-1}$, the probability of either of the two rankings two permutations removed is $r^3 (r + 2) [(1 + r)(1 + r + r^2)^2]^{-1}$, while the inverse ranking has probability $r^3 [(1 + r)(1 + r + r^2)]^{-1}$ of being the estimate. There is little gain over using only one process to estimate the standard ranking. For then the probability of correct and inverse ranking is the same as above for $m = 2$ while that of either ranking one permutation removed is $r(1 + r)^{-1}(1 + r + r^2)^{-1}$. This peculiarity seems to arise in large part, however, from the excessive number of ties possible in this special case.

In the estimation of p, the methods of the previous section are also relevant. That is, a consistent estimate of p as n becomes large is given by

$$\hat{p} = \frac{mn + s - 3}{mn + 2s - 3}, \quad \text{when} \quad \frac{s}{n(n-1)} < \frac{m}{4}, \tag{4.3}$$

and an appropriate modification is made otherwise. When n is small, however, it seems reasonable to consider only the unbiased estimates (3.7), to compute this estimate for each process, and then to average. Clearly, such an estimate is consistent in m and unbiased.

In the cases considered so far the association between the processes arose because there was a common standard ranking. The apparent differ-

ences in the observed rankings arose from fluctuations around this common standard. The number of permutations from the estimated standard ranking measured the tendency for the observed rankings to agree.

The comparison of several processes which do not have the same error or processes which do not have the same standard ranking involve more difficult problems which will not be treated in this paper. As rough practical guides, however, weighting the observed rankings in order to get better estimates of the standard order would in the first problem be an improvement over the unweighted scheme used earlier. Each ranking would lead to its own estimate of error.

If the standard ranking is not the same for all m processes, this would lead to greater disagreement among the observed rankings than if there was a common standard. The pooled estimate of p would then imply greater variability among the separate values of s/n for each process than is the case. One might, therefore, reject the hypothesis of common standard ranking when $\sum_{i=1}^{m} (s_{i\,ki} - \bar{s})^2/\mathrm{var}(s_i)$ is too small, using a chi square distribution with $m - 1$ degrees of freedom as a first approximation.

REFERENCES

[1] Feller, W. An introduction to probability theory and its applications, Vol. I. New York: Wiley, 1950.

[2] Kendall, M. G. The advanced theory of statistics, Vol. I, Chap. 16. London: Griffin, 1945.

[3] Blumen, I., Kogan, M., and McCarthy, P. J. The industrial mobility of labor as a probability process, Chap. 7. Ithaca, N. Y.: Cornell University, 1955.

Manuscript received 4/24/56

Revised manuscript received 8/27/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

ESTIMATION OF ERROR VARIANCES OF TEST SCORES

JOHN A. KEATS*

EDUCATIONAL TESTING SERVICE

The representation of test scores as n-dimensional points leads directly to an estimate of error variance at a particular score level in the case of equivalent items. Approximations are suggested for the case of non-equivalent items. These approximations are compared, with satisfactory results, with empirical data prepared by Dr. Mollenkopf.

General Representation of Test Scores

It is convenient for present purposes to consider the possible responses of a person to a test of n items as points in a space of n dimensions,

$$A = \{a_1\ a_2\ a_3\ \cdots\ a_n\},$$

when $a_i = 0$ *or* $a_i = 1$. If B is another point, i.e.,

$$B = \{b_1\ b_2\ b_3\ \cdots\ b_n\},$$

then define distance from A to B, i.e., $D_{(A,B)}$ as

$$D^2_{(A,B)} = [(a_1 - b_1)^2 + (a_2 - b_2)^2 \cdots (a_n - b_n)^2] = D^2_{(B,A)}\ .$$

In particular, the score S_A corresponding to A is given by $S_A = D^2_{(A,0)} = D^2_{(0,A)}$ = the squares of the distance of A from the origin.

It is also convenient to consider the arrangement in the n-dimensional space of points having the same score value. These arrangements will form regular $(n - 1)$-dimensional figures with $\binom{n}{s}$ vertices and with centers at the point

$$C_s = \left\{\frac{s}{n}\frac{s}{n}\frac{s}{n}\cdots\frac{s}{n}\right\}.$$

Since all patterns with the same score are commonly treated as equivalent, and since C_s is a type of average of such patterns, it is interesting to note that $D_{(C_s,0)} = s/\sqrt{n}$. This follows from the fact that $D^2_{(C_s,0)}$ = sum of squares of coordinates of $C_s = n(s^2/n^2) = s^2/n$, i. e., the score values corresponding to the centers of these figures mark off equal distances in the possible range 0 to n.

The next step is to consider the probability of occurrence of patterns. If all patterns are equally probable, then the distribution of scores will be

* Now at the University of Queensland, Brisbane, Queensland, Australia.

the binomial distribution $P(s) = \binom{n}{s}2^{-n}$. This case is trivial, as it corresponds to zero reliability. Of more importance is the case of equal probability of patterns for given score values, with no other restriction on the total probability for a particular score. In this case all items are equivalent statistically. This case will be examined further with respect to a problem raised by Mollenkopf [4].

The Variation of the Standard Error of Measurement

In his 1949 article Mollenkopf [4] studied the variation of the standard error of measurement with test score by considering the variance of the scores on a half test for persons who have the same score on the total test. Mollenkopf's demonstration [4, p. 191] that this variance is in fact related to the average error variance for persons with the same raw score is inadequate, as is Rulon's [6]. In both cases the inadequacy arises from the fact that the demonstrations deal with the average error variance irrespective of raw score. However if error variance varies with raw score, what is true on the average for all raw scores may not be true for each individual raw score.

Let s_j be the score of individual j, and the scores on two parallel halves of the test be s_{j1} and s_{j2} . Then

$$\begin{aligned} s_j &= s_{j1} + s_{j2} \\ &= t_{j1} + e_{j1} + t_{j2} + e_{j2} , \end{aligned}$$

where t stands for true score and e for error.

If individual j is retested many times with parallel tests,

$$\text{var } (s_j) = \text{var } (e_{j1} + e_{j2}),$$

since $t_{j1} = t_{j2} =$ constant.

This is equal to error variance in the usual sense. But

$$\text{var } (s_{j1} + s_{j2}) = 2 \text{ var } (s_{j1}).$$

Var (s_{j1}) may be estimated with one degree of freedom from one application of the test for each person. These variances may be averaged over persons at the same score level. This procedure leads to Mollenkopf's estimate of error variance, which is thus an estimate of error variance in the usual sense.

Consider now the case of equal probability of patterns for given score values. If the $\binom{n}{s}$ points corresponding to score s are further ordered in terms of their distance from the point $P = (P_1P_2 \cdots P_{n/2} \cdots P_n)$, where P_1 to $P_{n/2}$ are all 1 (n is taken as even) and $P_{(n/2)+1}$ to P_n are all 0, i.e., $s_P = n/2$, then the frequency distribution of the $\binom{n}{s}$ points in terms of the number (r)

of unities in the first $n/2$ coordinates will be

$$\frac{\binom{n/2}{r}\binom{n/2}{s-r}}{\binom{n}{s}} \quad \text{if} \quad s < n/2, \quad \text{and} \quad \frac{\binom{n/2}{r}\binom{n/2}{n-s-r}}{\binom{n}{n-s}} \quad \text{if} \quad s > n/2.$$

Furthermore, this will be a conditional distribution in the bivariate distribution obtained by plotting the score r on the first $n/2$ items against score $s - r$ on the second $n/2$ items. The condition is that total score is kept constant. The proof of these statements is given below.

Both of the distributions given have variances [1, p. 127]

$$\frac{1}{4}\frac{s(n-s)}{n-1}, \tag{1}$$

and, of course, this must equal

$$\frac{\sum_i (s_{i1} - s_{i2})^2}{4N_s}, \tag{2}$$

where $s_{i1} + s_{i2} = s$, and N_s is the frequency of score s, since $\sum_i (s_{i1} - s_{i2})^2 = \sum_i (2s_{i1} - s)^2 = 4 \sum_i (r_i - s/2)^2$ and since the items are equivalent $\bar{r}_j = s/2$. Hence (2) will equal (1) when N_s is arbitrarily large. The relation which Mollenkopf wanted to obtain was

$$\frac{\sum_i (s_{i1} - s_{i2})^2}{N_s} \tag{3}$$

as a function of $s = s_1 + s_2$. This is provided by the formula

$$\text{error variance} = \frac{s(n-s)}{(n-1)} \tag{4}$$

for the case of equivalent items.

Proof of Distribution

Consider

$$P = (P_1 = 1, P_2 = 1, P_3 = 1, \cdots P_{n/2} = 1, P_{n/2+1} = 0, \cdots P_n = 0)$$

and patterns for score $s < n/2$. Patterns which have the same number (r) of ones in the first $n/2$ elements will clearly have the same number $(s - r)$ of ones in the second $n/2$ elements and thus be the same distance from P. There will be

$$\binom{n/2}{r}\cdot\binom{n/2}{s-r}$$

such patterns and thus the probability of a pattern with score s falling into this rth category is

$$P(r) = \frac{\binom{n/2}{r}\binom{n/2}{s-r}}{\binom{n}{s}},$$

since all such patterns are assumed equally probable.

Clearly this rth category contains all patterns corresponding to a score r on the first half of the test and a score $s - r$ on the second half and so represents a cell in the bivariate distribution of the first half of the test against the second. The distribution $P(r)$ is the conditional distribution obtained by holding total score constant.

A similar proof holds for $s > n/2$, e.g., by considering the number of zeros instead of the number of ones.

Properties of the Formula (4) *for Error Variance*

From (3) and (4)

$$Y = \frac{s(n-s)}{n-1} = \frac{\sum_i (s_{i1} - s_{i2})^2}{N_s}.$$

To obtain the average error variance $\bar{Y}$ it is necessary to multiply by N_s, sum with respect to s, and divide by the total number of people (N). Then

$$\bar{Y} = \frac{1}{N(n-1)} \sum_{s=1}^{n} N_s s(n-s) = \frac{1}{N} \sum_{s=1}^{n} \sum_i (s_{i1} - s_{i2})^2$$

$$= \frac{n\bar{s} - \sigma_s^2 - \bar{s}^2}{n-1} = 2\sigma_{s1}^2(1 - r_{12}).$$

The extreme right-hand side can be shown to equal $\sigma_s^2(1 - r)$ when r is the correlation between the total test and a parallel test of the same length, i.e.,

$$\bar{Y} = \frac{\bar{s}(n-\bar{s})}{(n-1)} - \frac{\sigma_s^2}{(n-1)} = \sigma_s^2(1-r),$$

from which

$$r = \frac{n}{n-1}\left(1 - \frac{\bar{s}(n-\bar{s})}{n\sigma_s^2}\right).$$

This result is the Kuder-Richardson Formula 21 [2].

Lord [3], in his treatment of the error obtained from sampling items, arrived at an estimate of variance due to this error equal to $s(n-s)/n$ if n is sufficiently large so that s/n is a good estimate of true score. The present formula can thus be regarded as a small sample estimate corresponding to his formula.

The formula

$$Y = \frac{s(n - s)}{n - 1}$$

indicates that the error variance at a particular score level can be evaluated without using the reliability coefficient. This is an important result as it demonstrates that although the reliability coefficient varies from population to population for the same test, the error variance at a given score level remains constant.

The Effect of Non-equivalence

Whereas the formula developed is exact under the conditions stated, no exact statement can be made in general. However, an obvious adjustment can be made in the formula so that its average value will equal the average error variance for some typical group. Thus for practical purposes

$$Y = \frac{(1 - R)s(n - s)}{(1 - K)(n - 1)} = k\,\frac{s(n - s)}{(n - 1)},$$

since

$$\bar{Y} = \frac{1 - R}{1 - K}\,\sigma^2(1 - K) = \sigma^2(1 - R),$$

where R = reliability coefficient for some typical group,

K = the estimate from Kuder-Richardson formula 21 for the same group.

This average correction of Y cannot be expected to give good estimates of error variance in cases of extreme variation in item difficulty or of considerable heterogeneity of items. Its value can best be assessed by a comparison with data such as that used by Mollenkopf [4] and [5]. The graphs (Figures 1–13 below) indicate the degree to which adequate representation is obtained. For a more adequate description of the data see [4] and [5].

The points (×) indicate the mean of the squared differences for five-unit intervals of raw score. The solid line is the curve obtained from the theoretical formula. The curved broken line (---) is the second-degree curve of best fit presented by Mollenkopf, and the straight broken line (—·—·—) is the straight line obtained from the assumption that error variance is constant at all points of raw score. For all figures the values of the mean raw score (M), the standard deviation of raw scores (σ), the number of items (N) and the corrected split-half reliability (R) are included for the data on which the curves are based.

Figures 1–4 are for multiple-choice data and the corrected formula seems to give a reasonable representation here. Mollenkopf's formula is "not nearly as good representation of the error trend as the best fitting second degree curve" [5, p. 5]. The corrected formula gives a progressively worse fit as variation of item difficulty increases, as shown by cases 5–13 which

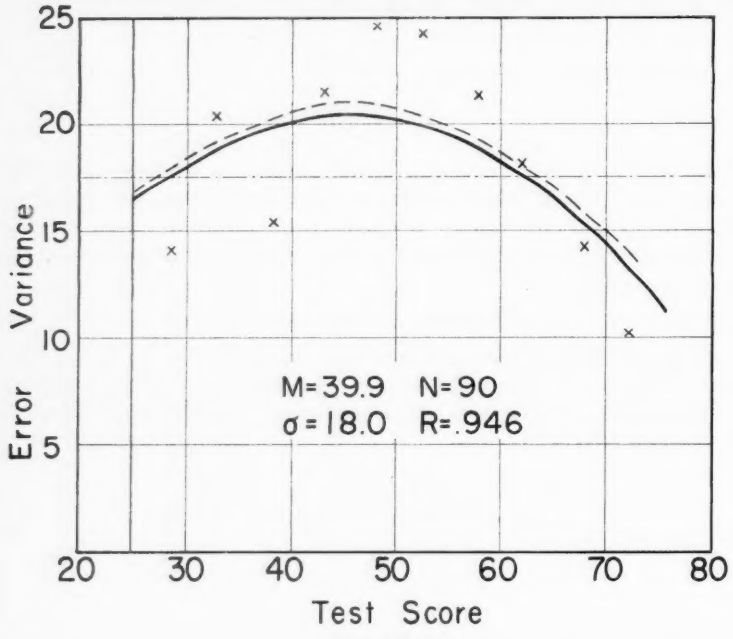

FIGURE 1

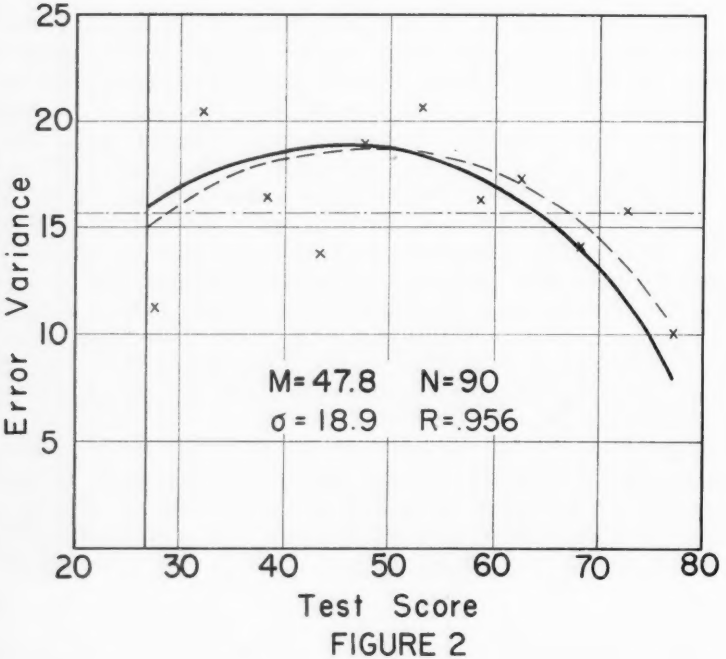

FIGURE 2

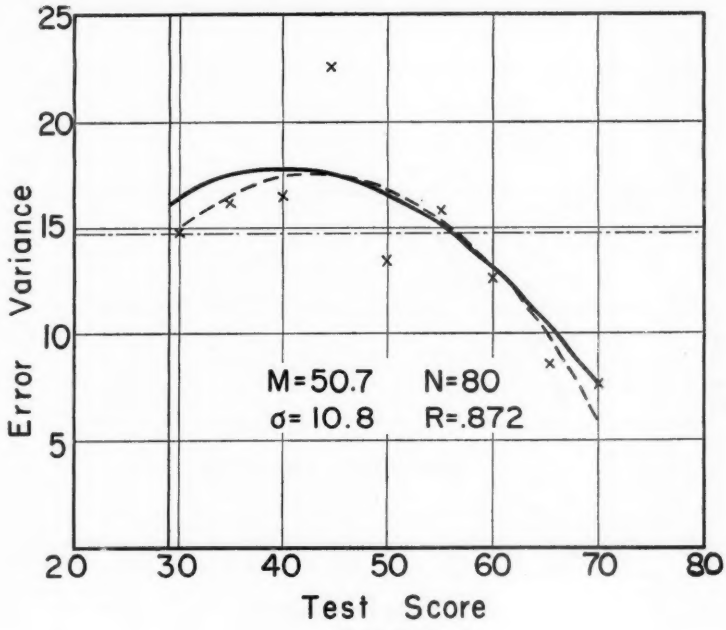

FIGURE 3

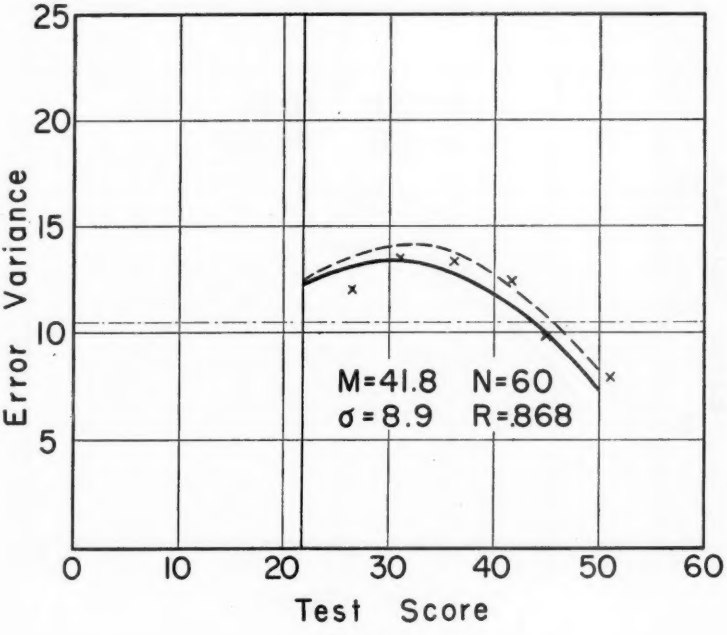

FIGURE 4

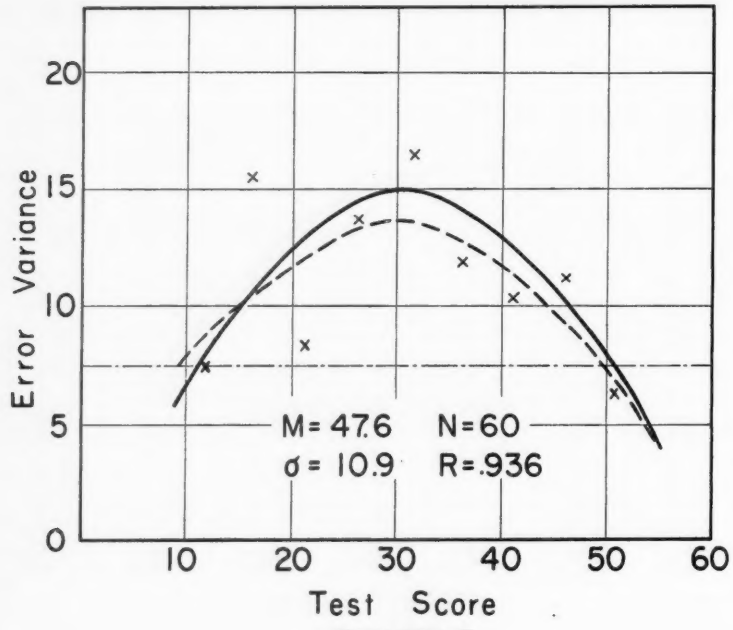

FIGURE 5

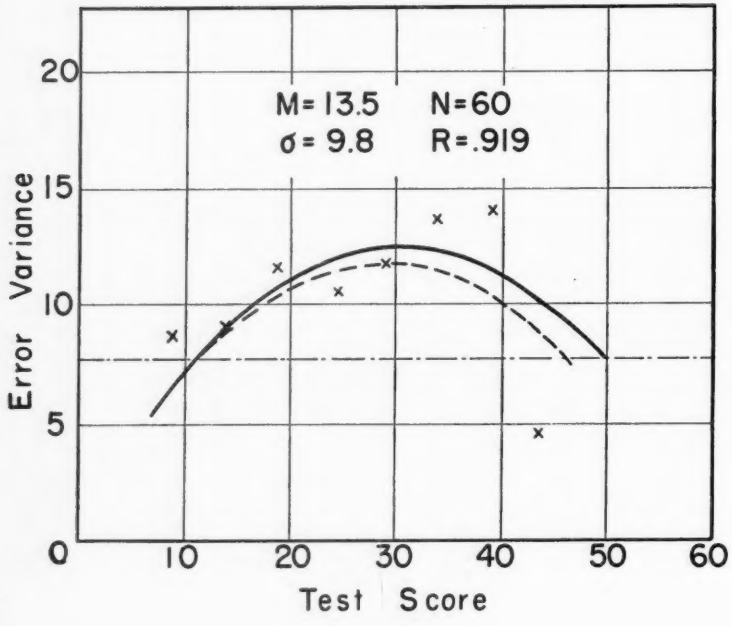

FIGURE 6

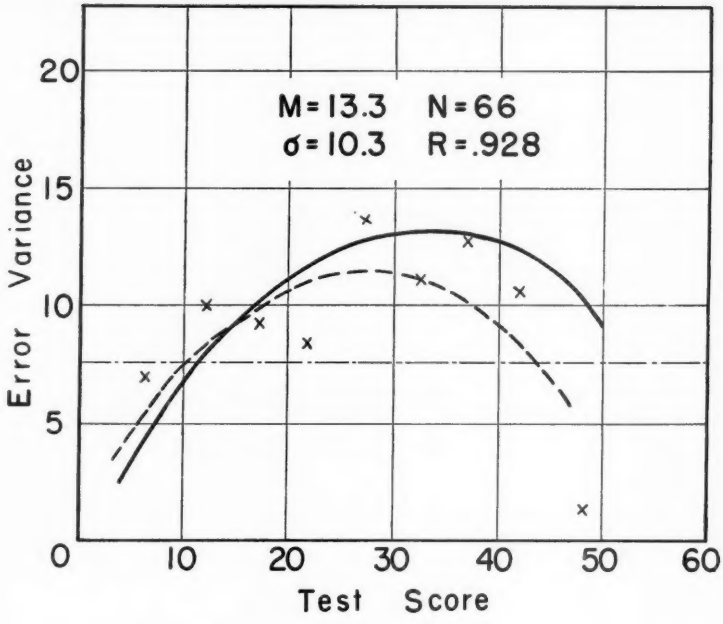

FIGURE 7

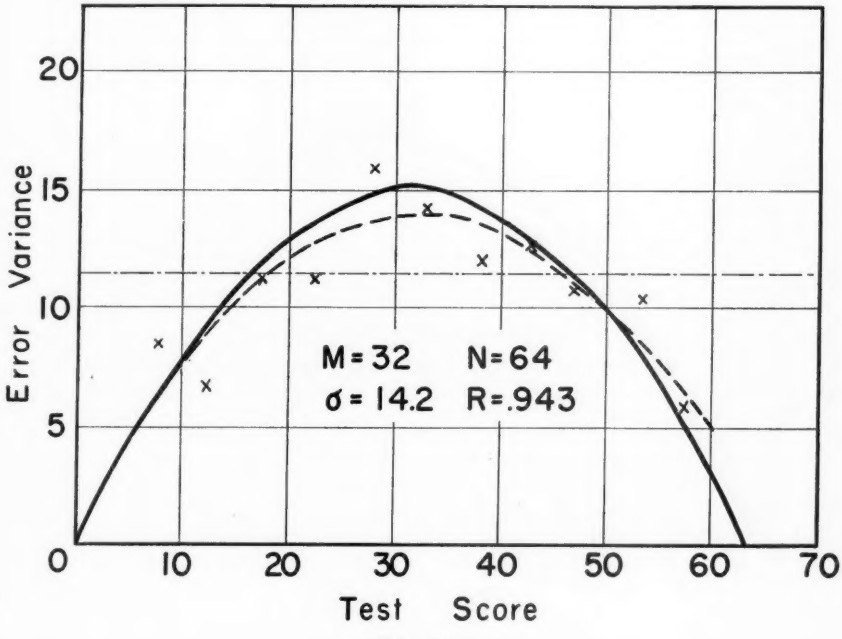

FIGURE 8

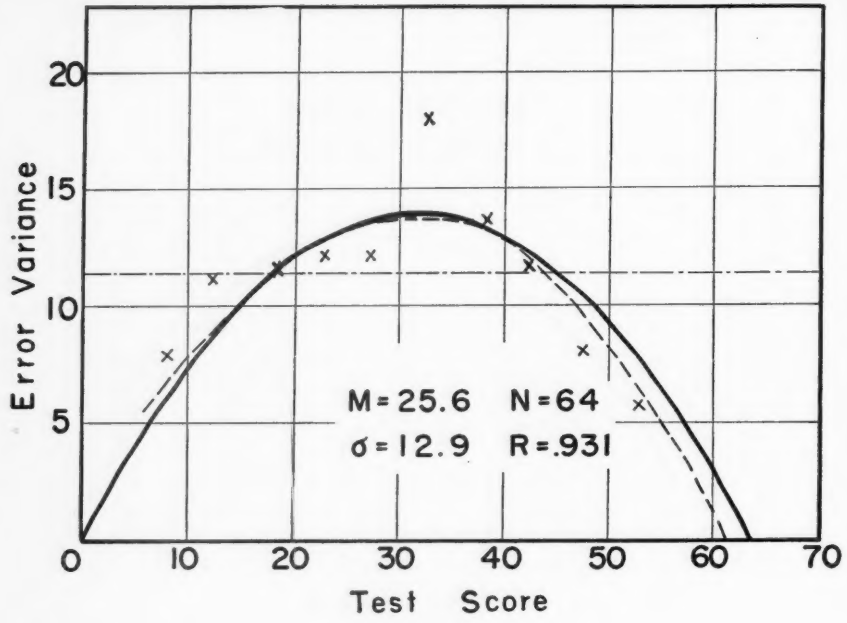

FIGURE 9

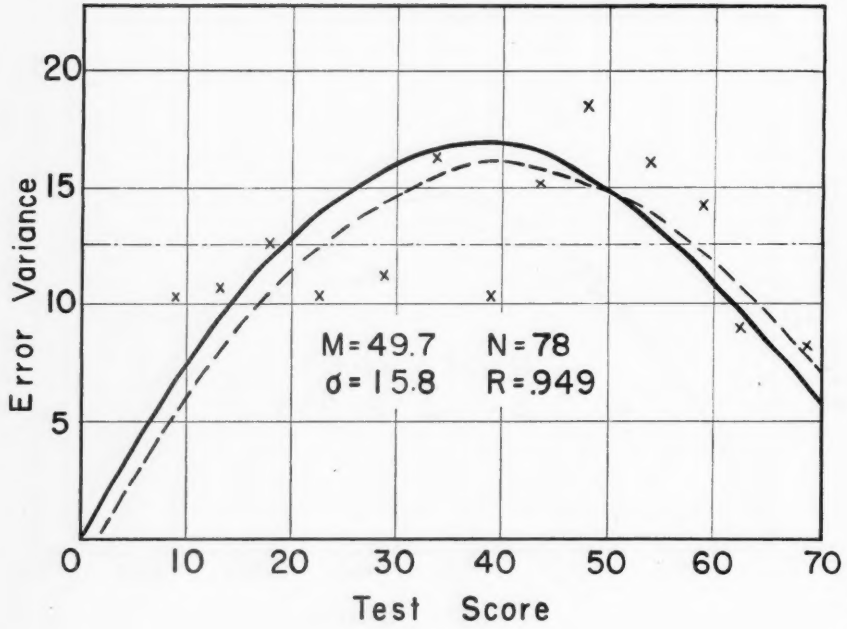

FIGURE 10

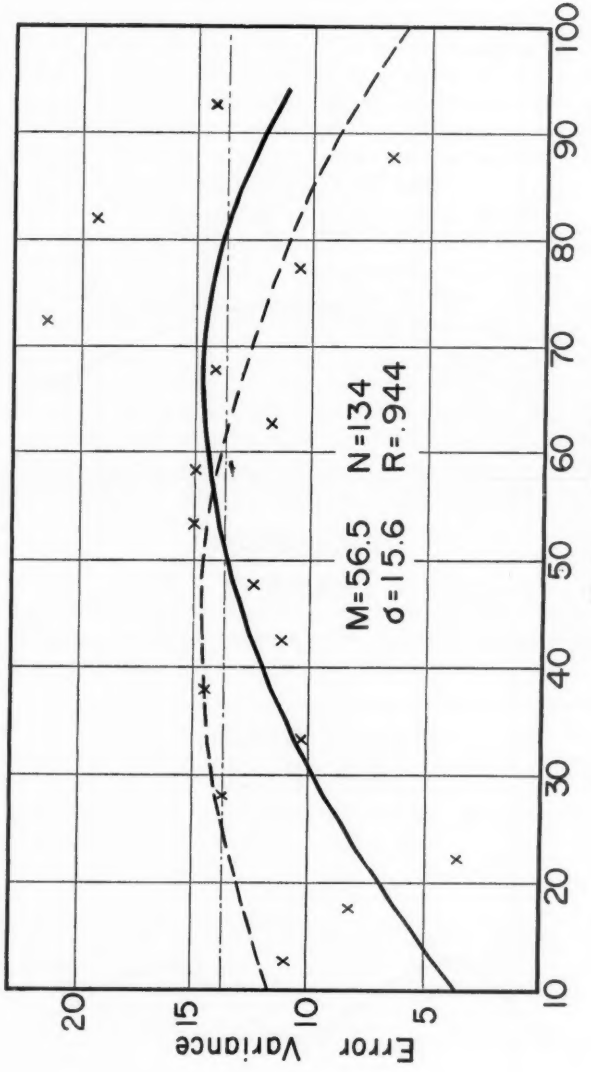
Error Variance
20
15
10
5
10
20
30
40
50
60
70
80
90
100
Test Score
M=56.5 N=134
σ=15.6 R=.944

FIGURE II

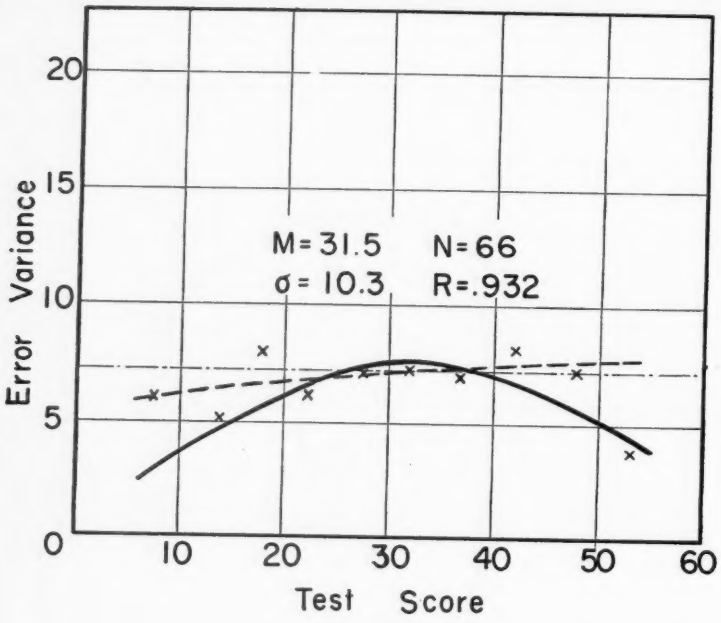

FIGURE 12

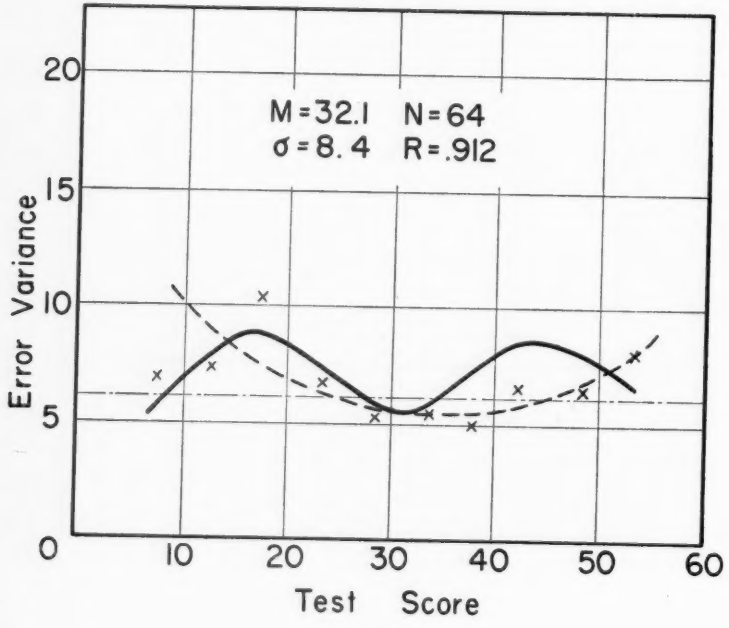

FIGURE 13

are arranged in order of increasing variation of item difficulty. The extent of variation of item difficulty is given in Table 6, page 218, of [4].

In case 13, in fact, the parabola has a minimum, not a maximum, and this could not be obtained by the corrected formula. The theoretical formulation given above could produce a minimum point by proceeding to a closer approximation to the general case, which leads to the formula

$$Y = \frac{n}{(n-1)}\left\{\frac{s(n-s)}{n} - n \text{ var } (p_s)\right\},$$

where var (p_s) is the variance of item difficulties for persons at score s. This formula was obtained by analogy with the formula for the variance of the distribution obtained by sampling from a binomial distribution with unequal probabilities as given by Kendall [1, p. 122]. It is not particularly useful as the computation of var (p_s) is tedious. Fortunately, very few tests used in practice have sufficient spread of item difficulty to require these computations. From the point of view of sampling of items this formula would be appropriate for stratified sampling, and the coarser approximation using k could be taken as an approximation to this case when the variance of item difficulties is small.

Applications

The expression $s(n-s)/(n-1)$ can always be used in the absence of group data if the assumptions made by Lord [3] in his estimates of error variance are considered reasonable, i.e., if the test is considered a result of random sampling from a population of items.

If group data are available, then k may be estimated and the formula used to estimate error variance in the usual sense at a particular score level. As k is used as a correction factor, it might be expected that it will be fairly stable across groups for the same test. Figures 1 and 2 refer to the same test given to two groups. It will be noticed that the two theoretical curves are more stable than the two values of $\sigma^2(1-R)$.

REFERENCES

[1] Kendall, M. G. The advanced theory of statistics. Vol. I. London: Griffin, 1947.
[2] Kuder, G. F. and Richardson, M. W. The theory of the estimation of test validity. *Psychometrika*, 1937, **2**, 151-160.
[3] Lord, F. M. Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, 1955, **20**, 1-22.
[4] Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189-229.
[5] Mollenkopf, W. G. Error of measurement variation for multiple choice tests. Educational Testing Service Research Bulletin 51-1, 1951.
[6] Rulon, P. J. Reliability of split-halves. *Harvard Educational Review*, 1939, **9**, 99-103.

Manuscript received 9/19/55
Revised manuscript received 5/7/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

THE DETAILED METHOD OF OPTIMAL REGIONS*

PAUL S. DWYER

UNIVERSITY OF MICHIGAN

The detailed method of optimal regions is an extended form of the method of optimal regions which has been found effective in solving the personnel classification problem when the number of job categories is small. The automatic determination of the successive values of the v_i, made possible by the more exact techniques of the detailed method, provide easier solutions for the more complex problems and provide solutions, which, for the most part, can be mechanized. In a sense the detailed method of optimal regions is more than a detailed form of the method of optimal regions. It is essentially a method of transformations by which the original matrix is reduced to a matrix from which the solution is easily obtained.

1. *Introduction*

The personnel classification problem [1] deals with the assignment of individuals to jobs, where the contribution to the common effort of each individual i if he is placed in position j is the known quantity, c_{ij} . Two recommended methods of solution are the simplex method [3] and the method of optimal regions [2]. The reader is referred to these references for the statement of the problem, the derivation of important properties, and descriptions of methods of solution.

The method of optimal regions is especially effective when, as is common, the number of different positions, k, is small. The method is based on the determination of a constant, v_j , for each position. In the detailed method of optimal regions, more specific rules are given for determining the v_j . Since these rules demand the calculation of auxiliary matrices, the detailed method is especially effective with machines, but it is also recommended when non-trivial problems are to be worked by hand.

Let the number of individuals to be assigned to the k positions be N, and let c_{ij} be entries in a matrix with N rows and k columns. The quota, q_j , the number of men to be assigned to position j, is exhibited in a row at the top of the matrix. This matrix is illustrated in Table 1, where ten men are to be assigned to four positions with quotas 4, 1, 4, 1, respectively. The problem is to make the assignment so that the sum of the corresponding c_{ij} values is as large as possible.

*Much of the basic research covered in this paper was carried out while the author was working on the problem of personnel classification in his capacity as Consultant, Personnel Research Branch, Department of the Army. The author wishes to express his appreciation to the Department of the Army for permission to use these materials in this paper. The opinions expressed are those of the author and are not to be construed as official or as those of the Department of the Army.

2. *The Conditions of Solution*

The basic conditions of solution, fundamental to the simplex method and other methods as well as to the method of optimal regions, imply the existence of u_i and v_j [2, p. 20] such that

$$c_{ij} = u_i + v_j \text{ for assigned values,} \tag{2.1}$$

$$c_{ij} \leq u_i + v_j \text{ for unassigned values.}$$

If J_i denotes the position to which individual i is assigned, the first expression of (2.1) is

$$c_{iJ_i} = u_i + v_{J_i} . \tag{2.2}$$

Subtracting the second expression of (2.1) from (2.2) gives a (necessary) condition for solution:

$$c_{iJ_i} - v_{J_i} \geq c_{ij} - v_j . \tag{2.3}$$

(2.3) may be called the generalized Brogden condition [2, pp. 20-21]. The method of optimal regions is based on the v_j of (2.3). The detailed method of optimal regions also uses the u_i of (2.2).

3. *The Determination of the Initial* v_j

Given the values c_{ij} and the quotas q_j , the first step of the detailed method of optimal regions (and of the method of optimal regions) is the determination of $v_j^{(0)}$, the initial values of v_j . Count out the q_j largest values in each column j and take the smallest of them. In Table 1, this process leads to the values, $v_1^{(0)} = 29$, $v_2^{(0)} = 49$, $v_3^{(0)} = 27$, $v_4^{(0)} = 41$. Then $c_{ij} - v_j^{(0)} \geq 0$ for at least q_j elements in column j.

In problems worked by hand, it is commonly useful to indicate those values which are equal to or greater than the $v_j^{(0)}$. Asterisks have been used to indicate those values.

The $v_j^{(0)}$ may be determined with the use of punched cards. One card is punched for each individual, indicating the c_{ij} values for all positions. The cards are then sorted for each position and the $v_j^{(0)}$ determined from the sorter card count or from a tabulator run using cumulated frequencies.

4. *The Determination of the Initial Assignment and the* $u_i^{(0)}$ *Values*

The initial assignment, $J_i^{(0)}$, is then made with the use of (2.3). Thus, in Table 1, compare $c_{ij} - v_j^{(0)}$ for successive values of j for each i and make the initial assignment $J_i^{(0)}$ to that column for which $c_{ij} - v_j^{(0)}$ is largest. Individual 1 is initially assigned to job category 1 since -6 is greater than -36, -11, or -27. In case of a tie for the largest value of $c_{ij} - v_j^{(0)}$, both values of j are recorded in the column for $J_i^{(0)}$.

With hand methods, many of the assignments can be made with a simpler rule. If there is a single asterisk in a given row, the assignment is made to the column in which the asterisk appears. If there are two or more asterisks in a given row, only those columns with asterisks need be considered in applying the criterion.

The number of initial assignments to each position is then determined. This number is indicated by $q_j^{(0)}$ and is placed, for comparison, above the q_j values. The first number indicates the number of definite initial assignments and the second the number of ties. Then form $q_j^{(0)} - q_j$, which indicates an excess of assignments if positive and a deficiency if negative. In Table 1 there is an excess of two assignments in column 1 and deficiencies of single assignments in columns 2 and 4.

Next determine the $u_i^{(0)}$ values. From (2.2)

$$u_i = c_{iJ_i} - v_{J_i}, \tag{4.1}$$

and then $u_i^{(0)} = c_{iJ_i}^{(0)} - v_{J_i}^{(0)}$. The results are placed in the column labelled $u_i^{(0)}$. In practice, it is commonly convenient to determine the values of u_i simultaneously with the values of J_i.

5. *The Determination of the First Transformed Matrix*

The first transformed matrix is computed using the formula

$$c_{ij}^{(1)} = c_{ij} - u_i^{(0)} - v_j^{(0)}. \tag{5.1}$$

Every element is either zero or negative since $u_i^{(0)}$ and $v_j^{(0)}$ are determined so that $c_{ij} \leq u_i^{(0)} + v_j^{(0)}$. The values $c_{ij}^{(1)}$ resulting from the application of (5.1) to the problem of Table 1 are shown in Table 2.

The negative signs in this matrix (and the following ones) can be eliminated by using the alternative transformation

$$t_{ij}^{(1)} = -c_{ij}^{(1)} = u_i^{(0)} + v_j^{(0)} - c_{ij}. \tag{5.2}$$

This is illustrated in Table 3. The value of $v_j^{(1)}$ is then the value of the q_jth *smallest* $t_{ij}^{(1)}$ in column j. The values of $J_i^{(1)}$ are then determined using

$$t_{iJ_i^{(1)}}^{(1)} - v_{J_i^{(1)}}^{(1)} \leq t_{ij}^{(1)} - v_j^{(1)}, \tag{5.3}$$

as illustrated in Table 3. The values of $J_i^{(0)}$ are indicated by the zero values of $t_{ij}^{(1)}$. The summary values $q_j^{(0)}$ and $q_j^{(1)}$ are recorded in the top rows. Examination shows that the transformation process is not yet completed since there is an excess of at least one assignment in position 1. Hence an additional transformation is carried out.

6. *The Determination of Successive Transformations*

Since the values of $v_j^{(1)}$ are available, only the values $u_i^{(1)}$ are needed to

complete the transformation. Now

$$u_i^{(1)} = c_{iJ_i^{(1)}}^{(1)} - v_{J_i^{(1)}}^{(1)}, \tag{6.1}$$

and the next transformation is given by

$$t_{ij}^{(2)} = t_{ij}^{(1)} - u_i^{(1)} - v_j^{(1)}. \tag{6.2}$$

The application of this transformation to the matrix of Table 3 leads to the matrix of Table 4. The symbol θ is used for each of the zero terms appearing in the same row. Thus the ties of Table 3 are indicated by the θ's of Table 4.

The values of $q_j^{(1)}$ show an excess of at least 1 in column 1. Hence one of the men tentatively assigned to column 1 must be assigned to one of the other columns. This is accomplished by subtracting from column 1 the smallest non-zero entry in any of the rows corresponding to individuals tentatively assigned to position 1. In Table 4, this value is 2; so $v_1^{(2)} = -2$. The remaining values of $v_j^{(2)}$ are 0, but they need not be recorded since nothing is to be subtracted.

The values of $J_i^{(2)}$ are then determined and the summary $q_j^{(2)}$ values. There are no excesses or deficiencies indicated either in the single columns or in the combinations of columns. The obvious assignment of ties leads to the set of J_i values identifying the solution.

In some cases it is necessary to make transformations on combinations of columns, since the method leads to a solution only when every combination of columns, as well as each column separately, has no deficiency [4, p. 16]. The technique for finding a suitable transformation when there is a deficiency in several columns differs slightly from that described above. In Table 3, note that an excess in column 1 indicates a deficiency in columns 2, 3 and 4. Indeed, a summary of the $J_i^{(1)}$ column shows only five men with 0 in columns 2, 3 or 4. But $q_2 + q_3 + q_4 = 6$. Hence there is a deficiency of 1 in this subset. A common positive amount can be subtracted from each of these columns to introduce an additional term, provided the negative of this amount is subtracted from every row which has at least one zero term in columns 2, 3, 4. In this way the tentative assignments to the columns having a net deficiency is maintained, while adding at least one new assignment to these columns. The amount to subtract from the columns is the smallest (non-zero) number in those columns which is not in a row tentatively assigned to column 2, column 3 or column 4. In this way the transformation leads to a matrix having the desired property that every element is non-negative.

In Table 4, $t_{43}^{(2)} = 2$; so $v_2^{(2)} = v_3^{(2)} = v_4^{(2)} = 2$ with $v_1^{(2)} = 0$. These values of $v_j^{(2)}$ lead to values of $J_i^{(2)}$ which are identical with those of Table 4. The two transformations are essentially equivalent transformations since they lead to the same matrix. This is the $t_{ij}^{(3)}$ matrix of Table 5. Assignments satisfying the quotas can be made to the zero terms of this matrix.

The method is designed, at each step, to decrease the number of deficiencies in some particular column or combination of columns without increasing the number of deficiencies in the remaining columns. The method necessarily converges since the total number of deficiencies is finite and since a sufficient condition for solution is an assignment with no deficiencies in any column or combination of columns [4, p. 16]. The process converges very rapidly in the common case in which the number of job categories is small. Experience has led to the empirical conclusion that, for small values of k, the number of transformations required for solution is approximately the value of k. Once the row deviates described in the next section are available, the number of transformations required is commonly less than $k/2$.

7. *The Use of Row Deviates*

A device which is useful in speeding the convergence of the method is the use of row deviates. Any constant may be subtracted from any row without changing the solution since (2.3) is not changed by subtracting a constant from c_{iJ_i} and from c_{ij}. Subtraction of the mean of the row from each element in the row results in row deviates from the mean. Preferably one may use large row deviates defined by

$$C_{ij} = k(c_{ij} - \bar{c}_i) = kc_{ij} - \sum_{j=1}^{k} c_{ij} = kc_{ij} - c_{i\cdot}\,, \tag{7.1}$$

where $c_{i\cdot}$ and $\bar{c}_i$ are, respectively, the sum and mean for row i.

The matrix of row deviates is then treated by the method described above. In the illustration used above the values of $U_i^{(0)}$ and $V_j^{(0)}$ obtained from the C_{ij} matrix are almost adequate for determining the solutions. This is shown in Table 6. Only a slight additional adjustment is necessary in column 4. The advantage of the use of the large row deviate transformation may be seen from the fact that the columns of the C_{ij} matrix are generally uncorrelated or slightly negatively correlated so that large values in one column are not apt to be accompanied by large values in some other column. The values of J_i in Table 6 are identical with those of Table 4.

8. *Solution of a Problem in the Frequency Form*

An illustration is next presented with $k = 5$ and in which it is necessary to analyze subsets of columns even though (large) deviates are used. For this purpose a frequency-form problem which Votaw and Dailey [4, p. 7] have worked with the simplex method is examined. A frequency-form problem results from the grouping of individual categories so that frequencies (f_i) as well as quotas (q_j) appear. The number of personnel categories is n.

The $n = 4$ values of f_i, as well as the $k = 5$ values of q_j, are shown in the first matrix of Table 7. The values of $c_{i\cdot}$ are first computed and then the

TABLE 1

Determination of $v_j^{(0)}$, $u_i^{(0)}$, $J_i^{(0)}$, and $q_j^{(0)}$

$q_j^{(0)}$	6-0	0-0	4-0	0-0			
i \ q_j	4	1	4	1	$u_i^{(0)}$	$J_i^{(0)}$	u_i
1	23	13	15	14	-6	1	-8
2	23	17	27*	37	0	3	0
3	30*	20	20	16	1	1	-1
4	26	22	22	15	-3	1	-5
5	28	25	30*	41*	3	3	3
6	29*	27	20	13	0	1	-2
7	29*	30	36*	38	9	3	9
8	25	33	20	22	-4	1	-6
9	25	37	33*	38	6	3	6
10	30*	49*	25	26	1	1	1
$v_j^{(0)}$	29	49	27	41	321		
v_j	31	48	27	38			315

TABLE 2

$c_{ij}^{(1)}$ Matrix

i \ q_j	4	1	4	1
1	0	-30	-5	-21
2	-6	-32	0	-4
3	0	-30	-8	-26
4	0	-24	-2	-23
5	-4	-27	0	-3
6	0	-22	-7	-28
7	-9	-28	0	-12
8	0	-12	-3	-15
9	-10	-18	0	-9
10	0	-1	-3	-16

TABLE 3

The $t_{ij}^{(1)}$ Matrix with $v_j^{(1)}$, $u_i^{(1)}$, $J_i^{(1)}$, and $q_j^{(1)}$

$q_j^{(1)}$	5-1	0-1	3-1	0-1			
$q_j^{(0)}$	6-0	0-0	4-0	0-0			
i \ q_j	4	1	4	1	$u_i^{(1)}$	$J_i^{(0)}$	$J_i^{(1)}$
1	0*	30	5	21	0	1	1
2	6	32	0*	4	0	3	3
3	0*	30	8	26	0	1	1
4	0*	24	2	23	0	1	1
5	4	27	0*	3*	0	3	3,4
6	0*	22	7	28	0	1	1
7	9	28	0*	12	0	3	3
8	0	12	3	15	0	1	1
9	10	18	0*	9	0	3	3
10	0*	1*	3	16	0	1	1,2
$v_j^{(1)}$	0	1	0	3	4		

TABLE 4

The $t_{ij}^{(2)}$ Matrix with $v_j^{(2)}$, $u_i^{(2)}$, $J_i^{(2)}$, $q_j^{(2)}$, and J_i

$q_j^{(2)}$	4-1	1-0	3-2	0-1				
$q_j^{(1)}$	5-1	0-1	3-1	0-1				
i \ q_j	4	1	4	1	$u_i^{(2)}$	$J_i^{(1)}$	$J_i^{(2)}$	J_i
1	0	29	5	18	2	1	1	1
2	6	31	0	1		3	3	3
3	0	29	8	23	2	1	1	1
4	0	23	2	20	2	1	1,3	3
5	4	26	0	0		3,4	3,4	4
6	0	21	7	25	2	1	1	1
7	9	27	0	9		3	3	3
8	0	11	3	12	2	1	1	1
9	10	17	0	6		3	3	3
10	0	0	3	13		1,2	2	2
$v_j^{(2)}$	-2				2			

TABLE 5

Final Matrix

$q_j^{(2)}$	4-1	1-0	3-2	0-1				
i \ q_j	4	1	4	1	$u_i^{(3)}$	$J_i^{(2)}$	J_i	c_{iJ_i}
1	0	27	3	16		1	1	23
2	8	31	0	1		3	3	27
3	0	27	0	21		1	1	30
4	0	21	0	18		1,3	3	22
5	6	26	0	0		3,4	4	41
6	0	19	5	23		1	1	29
7	11	27	0	9		3	3	36
8	0	9	1	10		1	1	25
9	12	17	0	6		3	3	33
10	2	0	3	13		2	2	49
$v_j^{(3)}$					315			315

values C_{ij} are recorded in the second matrix. In determining the values $v_j^{(0)}$ consider the frequencies associated with each row. Thus $V_1^{(0)} = -1$, since the $12 + 23$ values of -1 in column 1 are more than ample for the quota of 15. The values of $J_i^{(0)}$ are then obtained with the generalized Brogden condition. It is at once apparent that the columnar quotas can be met individually but that there is a deficiency in the subset of columns 1, 2, 5, since the $12 + 23 = 35$ men available cannot fill the $15 + 20 + 12 = 47$ jobs. A transformation is in order.

The values, $U_i^{(0)} = 0$, are computed and then the values $T_{ij}^{(1)}$ appearing in the third matrix. The values of $J_i^{(0)}$ summarize the zero terms. The deficiency in the subset consisting of columns 1, 2, 5 can be met after the matrix is reduced by subtracting some quantity from each of these columns to admit more zeros in the columns. The quantity to be subtracted is the smallest non-zero quantity in the rows not tentatively assigned to columns 1, 2, or 5. This quantity is 1; so $V_1^{(1)} = V_2^{(1)} = V_5^{(1)} = 1$ and, of course, $V_3^{(1)} = V_4^{(1)} = 0$.

The values $J_i^{(1)}$ are then determined. The number of available assignments in each row is so large that assignments satisfying the frequencies and quotas can be met in many different ways.

The additional transformation indicated by the values of $V_j^{(1)}$ is made so that the $T_{ij}^{(2)}$ matrix results. This transformation is not necessary to the solution, since a solution can be obtained from the last column of the third matrix, but the solution may also be obtained by making assignments to the zero terms of the last matrix in any way so as to satisfy the quotas and frequencies.

9. *The Determination of u_i and v_j*

It is now possible to determine the values of u_i and v_j of (2.1). If $t_{ij} = t_{ij}^{(m)}$ represents the final transform, let

$$\begin{aligned} t_{ij} &= 0 \text{ for assigned values,} \\ t_{ij} &\geq 0 \text{ for unassigned values.} \end{aligned} \tag{9.1}$$

Consider first the case in which the transformations are applied to the c_{ij} matrix without using row deviates. Then

$$t_{ij} = u_i^{(0)} + v_j^{(0)} - c_{ij} - (u_i^{(1)} + v_j^{(1)} + \cdots + u_i^{(m)} + v_j^{(m)}). \tag{9.2}$$

$$\begin{aligned} u_i &= u_i^{(0)} - u_i^{(1)} - u_i^{(2)} - \cdots - u_i^{(m)}, \\ v_j &= v_j^{(0)} - v_j^{(1)} - v_j^{(2)} - \cdots - v_j^{(m)}. \end{aligned} \tag{9.3}$$

The values of u_i and v_j for the problem of Table 1 were computed using (9.3) and are shown in the last column and row of Table 1.

The determination of u_i and v_j for problems using the large row deviate transformation is more involved. If the u_i and v_j appropriate to the C_{ij}

TABLE 6

Solution with Large Row Deviates

	c_{ij}					C_{ij}					
					$q_j^{(0)}$	3-1	1-0	5-1	0-0		
i \ q_j	4	1	4	1	$c_{i.}$ \ q_j	4	1	4	1	$U_i^{(0)}$	$J_1^{(0)}$
1	23	13	16	14	66	26*	-14	-2	-10	7	1
2	23	17	27	37	104	-12	-36	4*	44*	5	3
3	30	20	20	16	86	34*	-6	-6	-22	15	1
4	26	22	22	15	85	19*	3	3*	-25	4	3
5	28	25	30	41	124	-12	-24	-4	40	-3	3
6	29	27	20	13	89	27*	19	-9	-37	8	1
7	29	30	36	38	133	-17	-13	11*	19	12	3
8	25	33	20	22	100	0	32	-20	-12	-19	1,3
9	25	37	33	38	133	-33	15	-1*	19	0	3
10	30	49	25	26	130	-10	66*	-30	-26	0	2
v_j	5	16 3/4	0	11	1050	19	66	-1	44	211	

	$T_{ij}^{(1)}$							
$q_j^{(1)}$	3-1	1-0	4-2	0-1				
$q_j^{(0)}$	3-1	1-0	5-1	0-0				
i \ q_j	4	1	4	1	$U_i^{(1)}$	$J_i(0)$	$J_i(1)$	J_i
1	0*	87	8	61		1	1	1
2	36	107	0*	5		3	3	3
3	0*	87	20	81		1	1	1
4	4	67	0*	73		3	3	3
5	28	87	0*	1*		3	3,4	4
6	0*	55	16	89		1	1	1
7	48	91	0*	37		3	3	3
8	0*	15	0*	37		1,3	1,3	1
9	52	51	0*	25		3	3	3
10	29	0*	29	70		2	2	2
$V_j^{(1)}$				1	1			
V_j	19	66	-1	43				

TABLE 7

Solution of a Problem in the Frequency Form

	c_{ij}					
f_i \ q_j	15	20	25	28	12	$c_{i.}$
12	7	8	6	7	8	36
30	6	5	8	6	7	32
35	5	4	5	7	6	27
23	8	9	8	7	9	41
						3280

	C_{ij}						
f_i \ q_j	15	20	25	28	12	$U_i^{(0)}$	$J_i^{(0)}$
12	-1*	4*	-6	-1	4*	0	1,2,5
30	-2	-7	8*	-2	3	0	3
35	-2	-7	-2	8*	3	0	4
23	-1*	4*	-1	-6	4*	0	1,2,5
$V_j^{(0)}$	-1	4	8	8	4	537	1,2,5

	$T_{ij}^{(1)}$							
f_i \ q_j	15	20	25	28	12	$U_i^{(1)}$	$J_i^{(0)}$	$J_i(1)$
12	0	0	14	9	0	-1	1,2,5	1,2,5
30	1	11	0	10	1		3	1,3,5
35	1	11	10	0	1		4	1,4,5
23	0	0	9	14	0	-1	1,2,5	1,2,5
$V_j^{(1)}$	1	1			1	12	1,2,5	

	$T_{ij}^{(2)}$						
f_i \ q_j	15	20	25	28	12		$J_1^{(1)}$
12	0	0	15	10	0		1,2,5
30	0	10	0	10	0		1,3,5
35	0	10	10	0	0		1,4,5
23	0	0	10	15	0		1,2,5
						761	

matrix are U_i and V_j , a set of non-negative values of v_j can be determined from

$$v_j = (V_j - V_{j_0})/k,$$

where V_{j_0} is the smallest V_j . Thus in Table 6, the values of V_j are 19, 66, -1, 43; so the values of v_j are 5, 16 3/4, 0, 11. Again, in Table 7, the values of V_j are -2, 3, 8, 8, 3; so the values of v_j are 0, 1, 2, 2, 1. Other sets of v_j can be obtained by adding constants.

10. *The Determination of the Assignment Sum*

The assignment sum can be determined by applying the assignments for each row to the original c_{ij} matrix. This is illustrated in Table 5; the values of c_{iJ_i} are listed for each i, and the sum is 315 units. An alternative method is based on the formula

$$\sum c_{iJ_i} = (\sum u_i^{(0)} + \sum q_i v_i^{(0)}) - (\sum u_i^{(1)} + \sum q_i v_i^{(1)}) - \cdots - (\sum u_i^{(m)} + \sum q_i v_i^{(m)}). \tag{10.1}$$

The values in parentheses are given in the lower right corner of the respective matrices. If a problem in the frequency form is used, the values of $\sum u_i^{(\ )}$ are replaced by $\sum f_i u_i^{(\ )}$. If large row deviates are used, the appropriate formula is

$$\sum c_{iJ_i} = \frac{1}{k}\{\sum c_{i.} + (\sum U_i^{(0)} + \sum q_i V_i^{(0)}) - (\sum U_i^{(1)} + \sum q_i V_i^{(1)}) - \cdots - (\sum U_i^{(m)} + \sum q_i V_i^{(m)})\}. \tag{10.2}$$

Thus in Table 6,

$$\sum c_{iJ_i} = \frac{1050 + 211 - 1}{4} = 315 \text{ units}.$$

11. *Interpretation of the Method*

In a sense the detailed method of optimal regions is more than a detailed form of the method of optimal regions. For the former, specific rules are given for determining the successive increments to the v_j . It is essentially a method of reduced matrices in which an original matrix is transformed to a reduced matrix from which the assignment can be determined from the zero terms. The method is especially effective, particularly when using large row deviates, in solving non-trivial personnel assignment problems with a small number of positions.

REFERENCES

[1] Brogden, H. E. An approach to the problem of differential prediction. *Psychometrika*, 1946, **11**, 139-154.

[2] Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. *Psychometrika*, 1954, **19**, 11-26.

[3] Votaw, D. F. Methods of solving some personnel-classification problems. *Psychometrika*, 1952, **17**, 255–266.

[4] Votaw, D. F., Jr. and Dailey, J. T. Assignment of personnel to jobs. Research Bulletin **52-24**, Air Training Command, Human Resources Research Center. Lackland Air Force Base, August, 1952.

Original manuscript received 3/18/55

Revised manuscript received 10/10/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

THE DEVELOPMENT OF HIERARCHICAL FACTOR SOLUTIONS*

JOHN SCHMID

AND

JOHN M. LEIMAN

AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER

Although simple structure has proved to be a valuable principle for rotation of axes in factor analysis, an oblique factor solution often tends to confound the resulting interpretation. A model is presented here which transforms the oblique factor solution so as to preserve simple structure and, in addition, to provide orthogonal reference axes. Furthermore, this model makes explicit the hierarchical ordering of factors above the first-order domain.

The purpose of this paper is to present a procedure for transforming an oblique factor analysis solution containing a hierarchy of higher-order factors into an orthogonal solution which not only preserves the desired interpretation characteristics of the oblique solution, but also discloses the hierarchical structuring of the variables.

Oblique simple structure was proposed by Thurstone as a factor model useful for psychological research because of the simplicity with which interpretation could be made from a set of linear components underlying a set of scores. His argument is convincing when consideration is given to his "box problem" [9, pp. 140–146] for the factor loadings readily identify the dimensions of the boxes. In many studies, correlations among the reference axes make interpretation of simple structure difficult or questionable. In such cases usual methods of transformation from oblique to orthogonal axes fail to clarify the nature of the underlying parameters because many of the vanishing factor loadings become non-vanishing, thereby destroying simple structure. If one is willing to disavow the principle of parsimony of common factors, one may employ the type of factor solution outlined in this paper. This solution not only furnishes simple structure on orthogonal reference axes, but also provides a more complete rationale of the structuring of psychological traits than that given by (i) a conventional oblique solution or, for

*Grateful acknowledgment is given to Dr. Lloyd G. Humphreys for his encouragement and valuable suggestions in the development of this task. This investigation was carried out under the Air Force Personnel and Training Research Center program in support of Project Nos. 7702 and 7950. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government.

that matter, (ii) a solution in which the number of common factors is equal to the rank of the reduced correlation matrix.

It seems reasonable to assume that psychological behavior may be conceived as functioning at different levels of complexity. That is, a complex behavior activity might be thought of as an assembly of progressively less complex levels of activity—each level may have semantic, psychological, or practical meaning. For example, Vernon [11, pp. 22–24] reports that the mental structure of a group of British Army and Navy recruits was examined with a battery of cognitive tests. As determined from the sign pattern of centroid factor loadings, one general factor was found to be present in all tests. This factor was designated as g. With the elimination of g, the battery could be fractionated into two main groups of tests: academic and practical. In turn, the academic factor could be broken into verbal, numerical, and educational factors; the practical factor could be broken into mechanical, spatial, and physical factors. This structuring of the tests into a hierarchy of factors has many recommendable features—it provides information about the classification of tests and the behaviors measured by them in varying orders of concurrence and dependence. Had this particular centroid solution been rotated to an oblique solution, the hierarchical ordering would have been lost or rendered uncertain.

Structuring of tests into a hierarchical pattern is not a new consideration. Holzinger's bi-factor solution is a special case in which one second-order factor overlays the first-order group factors. Burt [1, 2, 3] has strongly advocated the hierarchical model for many years. His group factor method, which yields this hierarchy, proceeds by successive grouping of variables according to their sign pattern in a centroid solution. The procedure set forth in this paper, however, is an elaboration of the procedure demonstrated by Thompson [8, pp. 297–302] and Thurstone [9, pp. 411–439]. It differs from Burt's not in the product but in the process. The hierarchical solution is shown to be a consequence of successively obtained higher-order factor solutions. A necessary condition is the existence of simple structure at each level. If oblique simple structure exists, it can be recast into a hierarchical pattern similar in kind to that which Vernon inferred from the centroid solution. It will be seen that the characteristics of simple structure are retained not only at the level of the first-order factors but also at all levels.

Mathematical Rationale

The mathematical rationale for the model outlined in the paper is derived from Tucker's [10] generalization of the fundamental factor theorem stated by Thurstone [9, p. 78]. This theorem states that a correlation matrix, R, may be decomposed into correlated common factors and unique factors.

$$R = P\phi P' + U^2, \tag{1}$$

where P represents the coordinates of the vector representation of the variables on oblique Cartesian reference axes or factors, ϕ represents the intercorrelations among the oblique reference axes, and U represents the unique factor coefficients. It is readily seen that if ϕ is the identity matrix, the fundamental factor theorem of Thurstone results.

A second theorem used in this development also stems from Tucker's article. He shows that if the intercorrelations among the factors, ϕ, can be decomposed as

$$\phi = HH', \tag{2}$$

then the oblique factors, P, may be transformed into orthogonal factors, F, according to the operation

$$PH = F. \tag{3}$$

That is, the coordinates of the variables represented as vectors may be transformed from oblique to orthogonal reference axes. Each row of H represents the direction cosines of the oblique axes with respect to the orthogonal axes developed by the decomposition.

Guttman [4] demonstrates that if a matrix of intercorrelations, ϕ, is factored as in (2) no matter how H is built up, the reference axes are orthogonal. The factoring or decomposition may involve any of a variety of procedures, such as the diagonal or square root method of factoring, the centroid procedure, or the method of principal axes.

The development of the hierarchical model utilizes these propositions. In the following discussion, P_i will refer to the primary factor pattern of the ith order variables or factors; that is, the coordinates of the vector representation of the variables on the ith order oblique reference axes. R_i will be used to designate the intercorrelations among the ith primary factor reference axes. U_i will represent the unique ith order variables or factors.

At the outset, the initial correlation matrix, R, is decomposed according to (1) as follows:

$$R = P_1R_1P_1' + U_1^2 . \tag{4}$$

In like manner, R_1 is decomposed

$$R_1 = P_2R_2P_2' + U_2^2 . \tag{5}$$

In turn, R_2 is decomposed

$$R_2 = P_3R_3P_3' + U_3^2 . \tag{6}$$

Each higher-level matrix of intercorrelations among primary factors is decomposed in this fashion until R_i becomes the identity matrix, which implies that the ith order primary factors are orthogonal. That is,

$$R_{i-1} = P_iP_i' + U_i^2 . \tag{7}$$

In many cases, R_i becomes a unit scalar and P_i, therefore, is merely a column matrix. Elementary matrix manipulation permits (7) to be rewritten as a product of a supermatrix and its transpose:

$$R_{i-1} = [P_i : U_i] \cdot [P_i : U_i]'. \tag{8}$$

Designating the supermatrix, $[P_i : U_i]$, by B_i, according to (3), the $(i-1)$th order primary factors, P_{i-1}, can be made orthogonal by the operation $P_{i-1}B_i$.

However

$$R_{i-2} = P_{i-1}R_{i-1}P'_{i-1} + U^2_{i-1}. \tag{9}$$

Therefore, it follows that R_{i-2} may be rewritten as a product of a new supermatrix and its transpose:

$$R_{i-2} = [P_{i-1}B_i : U_{i-1}] \cdot [P_{i-1}B_i : U_{i-1}]'. \tag{10}$$

This new supermatrix may be designated as B_{i-1}. By virtue of (2) and Guttman's demonstration [4], orthogonal reference axes are obtained. Furthermore, B_{i-1} serves to rotate the primary pattern, P_{i-2}, to this orthogonal reference framework. Continuing this process to the lowest-order level, the initial primary or first-order factors, P_1, are orthogonalized by the operation P_1B_2. Designate P_1B_2 as B instead of B_1 since one is usually not concerned with explicitly appending the diagonal matrix of unique factors to the common factor solution. B, then, is the hierarchical solution. Since

$$R \text{ (with communalities)} = BB', \tag{11}$$

B represents coordinates of the test variables on orthogonal axes.

In the development of a hierarchical solution, careful attention should be paid to the distinction between simple structure and primary pattern. This distinction has been clearly drawn and illustrated by Harris and Knoell [5]. The hierarchical solution is contingent upon the development of a primary pattern at each level. This primary pattern, however, may be obtained from the simple structure, which is computed either graphically or analytically. Once simple structure is identified, it may easily be converted to primary pattern [5] by the operation

$$P_i = V_i(R_i^{-1})_d^{\frac{1}{2}}, \tag{12}$$

where P_i is primary pattern, V_i is simple structure, and $(R_i^{-1})_d^{\frac{1}{2}}$ is the matrix of the reciprocals of the direction cosines between each primary axis and its own simple structure reference axis. $(R_i^{-1})_d^{\frac{1}{2}}$ is obtained by taking the square roots of the diagonal elements only of R_i^{-1}.

Procedure

To demonstrate the procedure for rotating an oblique simple structure into a hierarchical factor solution, a correlation model was constructed from

TABLE 1

Correlation Matrix, R*

	1	2	3	4	5	6	7	8	9	10	11	12
1	6400	7200	3136	2688	0983	0491	1290	0369	2903	1613	0645	0753
2		8100	3528	3024	1106	0553	1452	0415	3266	1814	0726	0847
3			4900	4200	0753	0377	0988	0282	2222	1235	0494	0576
4				3600	0645	0323	0847	0242	1905	1058	0424	0494
5					6400	3200	1344	0384	1089	0605	0242	0282
6						1600	0672	0192	0544	0302	0121	0141
7							4900	1400	1429	0794	0318	0370
8								0400	0408	0227	0091	0106
9									8100	4500	1458	1701
10										2500	0810	0945
11											3600	4200
12												4900

*Communalities appear in the principal diagonal. Decimal points have been omitted.

TABLE 2

Primary Pattern, P_1

	I	II	III	IV	V	VI
1	.8					
2	.9					
3		.7				
4		.6				
5			.8			
6			.4			
7				.7		
8				.2		
9					.9	
10					.5	
11						.6
12						.7

TABLE 3

Intercorrelations of Primary Factors, R_1

	I	II	III	IV	V	VI
I	1.0000	.5600	.1536	.2304	.4032	.1344
II	.5600	1.0000	.1344	.2016	.3528	.1176
III	.1536	.1344	1.0000	.2400	.1512	.0504
IV	.2304	.2016	.2400	1.0000	.2268	.0756
V	.4032	.3528	.1512	.2268	1.0000	.2700
VI	.1344	.1176	.0504	.0756	.2700	1.0000

TABLE 4

Second-Order Primary Factors, P_2

	I	II	III
1	.8		
2	.7		
3		.4	
4		.6	
5			.9
6			.3

TABLE 5

Correlations Among Second-Order Primary Factors, R_2

	I	II	III
I	1.0000	.4800	.5600
II	.4800	1.0000	.4200
III	.5600	.4200	1.0000

a postulated simple structure factor matrix. It should be emphasized, however, that any set of empirical variables which can be rotated to simple structure can also be put in this more interpretable and meaningful hierarchical form. That is, if simple structure exists by any definition for a set of variables, the procedure is applicable. The given correlation matrix is presented in Table 1. An oblique solution was developed by the multiple-group method [6]. This oblique solution consists of a primary pattern, P_1 , and intercorrelations among the primary factors, R_1 . These two matrices are presented in Tables 2 and 3. An oblique solution could have been produced by rotation from a centroid solution or by some analytic method instead of the multiple-group procedure. However, the method of arriving at the oblique solution is of little consequence for our purposes, and the grouping procedure was thought to be the most expeditious here. If oblique simple structure, V_1 , had been produced, P_1 could be obtained quite readily by the operation indicated in (12). Regardless of methodology, the final rotated oblique solution should be transformed into a primary pattern, P_1 , as defined by Holzinger and Harman [7, chap. XI].

The intercorrelations of the primary factors, R_1 , (with communalities determined and placed in the diagonal elements) are then factored by any method. Usually it is most expeditious to carry out a *common*-factor analysis at each stage to separate the common-factor space from the unique-factor space. Rotation of these second-order factors is then performed to obtain the primary pattern of the second-order factors, P_2 , (Table 4) and the intercorrelations of the second-order primary factors, R_2 , (Table 5). A check may be made at this point since R_1 (with communalities) $= P_2R_2P_2'$. Again this P_2 may be developed by the construction of an oblique simple structure, V_2 , which is then transformed into P_2 by the operation indicated in (12).

Since the second-order factors, P_2 , are correlated, it is obvious that a third-order factor exists. Consequently, R_2 is factored. Factoring shows that there is one third-order factor and three unique factors, B_3 (see Table 6). The progressive factoring of higher orders is now complete. This information is used for developing the preferred hierarchical factor solution. To do this, the operation P_2B_3 is performed (Table 7) and the matrix of unique factors of R_1 , U_2 , is appended as shown in Table 8. That is, $B_2 = [P_2B_3 : U_2]$. It should be noted that $B_2B_2' = R_1$ (with unities in the diagonal of R_1). This matrix, B_2 , is used as the transformation matrix for rotating the first-order oblique solution, P_1 , into the final hierarchical solution, B (Table 9), according to the operation

$$B = P_1B_2 . \tag{13}$$

This procedure may be extended to higher orders if correlations are found among fourth-order or higher-order factors.

It will be observed that this hierarchical solution contains 10 common

TABLE 6

Third-Order Common and Unique Factors, B_3

	P_3	U_3		
	I	II	III	IV
1	.8000	.6000		
2	.6000		.8000	
3	.7000			.7141

TABLE 7

Orthogonalized Second-Order Common Factors, P_2B_3

	I	II	III	IV
1	.6400	.4800		
2	.5600	.4200		
3	.2400		.3200	
4	.3600		.4800	
5	.6300			.6427
6	.2100			.2142

TABLE 8

Orthogonalized Second-Order Common and Unique Factors, B_2

	P_2B_3				U_2					
	I	II	III	IV	V	VI	VII	VIII	IX	X
1	.6400	.4800			.6000					
2	.5600	.4200				.7141				
3	.2400		.3200				.9165			
4	.3600		.4800					.8000		
5	.6300			.6427					.4359	
6	.2100			.2142						.9539

TABLE 9

Hierarchical Factor Solution, B

	I	II	III	IV	V	VI	VII	VIII	IX	X
1	.5120	.3840			.4800					
2	.5760	.4320			.5400					
3	.3920	.2940				.4999				
4	.3360	.2520				.4285				
5	.1920		.2560				.7332			
6	.0960		.1280				.3666			
7	.2520		.3360					.5600		
8	.0720		.0960					.1600		
9	.5670			.5784					.3923	
10	.3150			.3214					.2180	
11	.1260			.1285						.5723
12	.1470			.1499						.6677

factors, where all tests define factor I. Factors II, III, and IV are the next most complex factors. Each of these in turn can be broken down into the finer composites illustrated by factors V through X. These last six factors identify the six factors of the original oblique solution, P_1. It will be observed that this solution reproduces the communalities and the off-diagonal correlations of the original correlation matrix exactly. Furthermore, it furnishes the same factorial interpretation as is found in the oblique solution, P_1, which is the usual type of solution obtained by researchers. Ease of psychological interpretation has not been sacrificed by the use of the hierarchical solution, and what was concealed in the intercorrelations of the oblique

axes now takes on added meaning in terms of the progressive groupings of the variables at higher levels.

It should be emphasized that even though the oblique solution, P_1, contains variables of complexity one only, this is not a restriction. Variables of any complexity may be used.

Discussion

A question arises about the stability of the hierarchical solution upon modification of the battery of tests. Burt concludes [3, p. 70] that the hierarchical solution—designated by him as the group-factor solution—remains "stable, if not absolutely invariant, even when the battery of tests or traits is modified, e.g., when a comparatively small battery is enlarged by the addition of more tests or more groups of tests, or when a large battery is curtailed by the omission of tests." The introduction of a new group of tests which are unrelated to any group already in the battery would, of course, add a new group factor.

In all probability, selection, univariate and multivariate, and sampling variation would affect this model in the same manner as the simple structure model. These points concerning battery modification, selection, and sampling stability need further research for clarification.

Practical applications of this model will be greatly aided as more objective and analytical criteria and techniques for transformation to simple structure are achieved. Nevertheless, even with present methods of attaining simple structure, the hierarchical solution is useful.

Summary of Steps as Applied to Illustration

1. R, with communalities, was factored into P_1 and R_1 (Tables 1, 2, and 3), that is

$$R \text{ (with communalities)} = P_1 R_1 P_1' .$$

2. R_1, with communalities, was factored into P_2 and R_2 (Tables 4 and 5), that is

$$R_1 \text{ (with communalities)} = P_2 R_2 P_2' ,$$

$$R_1 \text{ (with unities)} = P_2 R_2 P_2' + U_2^2 ,$$

where U_2 represents the diagonal matrix of unique factors of R_1.

3. R_2, with communalities, was factored into P_3. (Table 6). One common factor was found, i.e. R_3 was a unit scalar.

$$R_2 \text{ (with communalities)} = P_3 P_3' ,$$

$$R_2 \text{ (with unities)} = P_3 P_3' + U_3^2 .$$

4. When only one common factor remains, as in this illustration, factoring

of the higher-order matrices is completed. Otherwise, the procedure would be continued until R_i becomes an identity matrix or a single highest-order factor is found. At this stage, these intermediate matrices are used for constructing a rotation matrix for transforming the primary pattern, P_1 , into a hierarchical solution, B.

5. Form matrix B_3 by appending the unique-factor loadings of R_2 to P_3 , that is

$$B_3 = [P_3 : U_3]. \text{ (Table 6).}$$

It follows that

$$R_2 \text{ (with communalities)} = P_3P_3' ,$$

$$R_2 \text{ (with unities)} = B_3B_3' .$$

6. Carry out the matrix operation P_2B_3 . (Table 7).

7. Form matrix B_2 by appending the unique-factor loadings of R_1 to P_2B_3 , that is

$$B_2 = [P_2B_3 : U_2]. \text{ (Table 8).}$$

It follows that

$$R_1 \text{ (with communalities)} = P_2B_3B_3'P_2' ,$$

$$R_1 \text{ (with unities)} = B_2B_2' .$$

8. The hierarchical solution, B, then is constructed by the operation

$$B = P_1B_2 . \text{ (Table 9).}$$

REFERENCES

[1] Burt, C. Alternative methods of factor analysis. *Brit. J. Psychol. (Statist. Sec.)*, 1949, **2**, 98-121.
[2] Burt, C. Subdivided factors. *Brit. J. Psychol. (Statist. Sec.)*, 1949, **2**, 41-63.
[3] Burt, C. Group factor analysis. *Brit. J. Psychol. (Statist. Sec.)*, 1950, **3**, 40-75.
[4] Guttman, L. General theory and methods for matric factoring. *Psychometrika*, 1944, **9**, 1-16.
[5] Harris, C. W. and Knoell, D. M. The oblique solution in factor analysis. *J. educ. Psychol.*, 1948, 385-403.
[6] Holzinger, K. J. A simple method of factor analysis. *Psychometrika*, 1944, **9**, 257-261.
[7] Holzinger, K. J. and Harman, H. H. Factor analysis. Chicago: Univ. Chicago Press, 1951.
[8] Thomson, G. H. The factorial analysis of human ability. New York: Houghton Mifflin, 1948.
[9] Thurstone, L. L. Multiple-factor analysis. Chicago: Univ. Chicago Press, 1947.
[10] Tucker, L. R. The role of correlated factors in factor analysis. *Psychometrika*, 1940, **5**, 141-152.
[11] Vernon, P. The structure of human abilities. New York: Wiley, 1950.

Manuscript received 1/26/56
Revised manuscript received 6/9/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

A THEORY OF PATTERN ANALYSIS FOR THE PREDICTION OF A QUANTITATIVE CRITERION

ARDIE LUBIN

WALTER REED ARMY INSTITUTE OF RESEARCH

AND

HOBART G. OSBURN

HUMAN RESOURCES RESEARCH OFFICE, GEORGE WASHINGTON UNIVERSITY

A method of pattern analysis is presented for the case of dichotomous items and a quantitative criterion. This "configural scale" has maximum validity in the least squares sense. A technique for computing the configural scale as a polynomial function of the item scores is given. Tests of significance are outlined for such questions as: Is there a linear or non-linear relation between the quantitative criterion and the item scores? Does the addition of certain items to the test increase the validity of the configural scale? Are all the items in the configural scale fully effective?

1. *Introduction*

The purpose of this paper is to derive an optimal method of pattern analysis for the prediction of a quantitative criterion. Suppose each individual has taken a test of t items, and each individual has a criterion score on a quantitative variable. What is the best possible way of predicting the criterion score from the individual's answer pattern? How can we make use of *all* the information given by the responses to the t items?

A least squares method will be proposed as an adequate solution for the case of a quantitative criterion. This method is satisfactory if the objective is to minimize the sum of squared errors when predicting from the subject's answer pattern to his criterion score.

Guttman [2] and Rao [8] have already noted that when the criterion is qualitative, the maximum likelihood solution will produce the minimum number of misclassifications. Rao has given a general proof of this, which holds whether the predictors are quantitative or qualitative. The least squares solution presented in this paper is equivalent to using maximum likelihood when the distribution of criterion scores within the answer pattern is normal.

Meehl [7] called attention to a special case where two dichotomous items which correlated zero with the dichotomous criterion would give a validity of unity when the item patterns were scored by what Meehl termed

the *configural* method. Meehl's configural scoring method can be derived from Rao's maximum likelihood approach. Horst [4] has shown that Meehl's configural scoring corresponds to a polynomial function of the dichotomous item scores.

2. *Definition of the Configural Scale*

Given a test of t dichotomous items, there will be 2^t possible answer patterns. If the mean criterion score is calculated for each answer pattern, there will be 2^t possible criterion means. Every individual in an answer pattern is assigned the same score, the mean criterion score for that answer pattern. This set of scores is called a *configural scale* (by analogy with Meehl's scoring method). It will be shown that an individual's configural score is, in the least squares sense, the best prediction of his criterion score.

3. *Theorems on the Configural Scale*

THEOREM 1. *The zero-order correlation of the configural scale with the criterion is equal to or greater than the correlation of the criterion with any other set of scores based on the answers to the t dichotomous items.*

In other words, the configural scale based on criterion means has maximum validity. The proof of this theorem follows from the least squares property of the mean. All individuals in an answer pattern have exactly the same item scores and cannot be distinguished from each other on the basis of the test. So they must all be assigned the same predicted criterion score. What score, when assigned to all individuals who have the same answer pattern, will produce the smallest sum of squared deviations from the observed criterion scores? This score is, of course, the answer-pattern mean.

The general Pearson correlation coefficient can be defined as

$$r = \sqrt{1 - (W/T)}, \tag{1}$$

where T equals the sum of squared deviations about the general criterion mean, and W equals the sum of squared deviations of the observed from the predicted criterion scores. Since W is a minimum when the configural scale is used, r must be a maximum.

This general least squares technique of predicting quantitative scores from qualitative attributes has been known for some time. Guttman [2] gave essentially the same method in his section on "The prediction of a quantitative variate from a set of attributes." He pointed out that the correlation in (1) is actually η, the correlation ratio. It is also equal to the product-moment correlation of the configural scores with the observed criterion scores.

The coefficient in (1) can be defined in analysis of variance terms. Let B equal the deviance (sum of squares) between the answer pattern means;

let W be the deviance (sum of squares) within answer patterns. Then

$$r = \sqrt{B/(B + W)}; \tag{2}$$

this formula is equivalent to (1). In this way configural scale analysis can be translated into R. A. Fisher's terminology of *between* groups and *within* groups.

To summarize, a configural scale has been defined as the set of 2^i criterion averages, one for each answer pattern, and has been shown to possess maximum validity in the least squares sense.

THEOREM 2. *The configural scale can be represented as a polynomial function of the item scores:*

$$\hat{C} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_iX_i + b_{12}X_1X_2 + b_{13}X_1X_3 + \cdots + b_{123}X_1X_2X_3 + \cdots . \tag{3}$$

As an example of Theorem 2, take the two-item configural scale set forth in Table 1 where $\bar{C}_i$ is the criterion mean of the ith answer pattern.

TABLE 1

A Two-item Configural Scale

Answer Pattern	Items		Criterion Average
	1	2	
1	yes	yes	$\bar{C}_1$
2	yes	no	$\bar{C}_2$
3	no	yes	$\bar{C}_3$
4	no	no	$\bar{C}_4$

The polynomial predictor is

$$\hat{C} = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 , \tag{4}$$

where X_1 is the score on item 1,
X_2 is the score on item 2,
$\hat{C}$ is the predicted criterion score, and
b_0 , b_1 , b_2 , and b_{12} are the best fitting regression coefficients.

In this paper it is arbitrarily assumed that a *No* response is scored zero and a *Yes* response is scored unity. So, for the *Yes-Yes* answer pattern: $X_1 = 1$, $X_2 = 1$, and therefore $X_1X_2 = 1$. It follows that $\hat{C}_1 = b_0 + b_1 + b_2 + b_{12}$. For the *Yes-No* answer pattern: $X_1 = 1$, $X_2 = 0$, and therefore $X_1X_2 = 0$. It follows that $\hat{C}_2 = b_0 + b_1$. In a similar way, equations which

involve only the unknown b's and the predicted $\hat{C}$'s can be derived for each of the answer patterns.

Note that X_1X_2 turns out to be a dichotomous score of either zero or unity. In general, all possible multiplicative combinations of the item scores will be either zero or unity.

There are four unknown coefficients in (4); there are 2^2 or four means in Table 1. Therefore, an exact solution for the unknown coefficients, such that $\hat{C} = \overline{C}$, is always possible. The four equations are

$$\overline{C}_1 = b_0 + b_1 + b_2 + b_{12} , \tag{5}$$

$$\overline{C}_2 = b_0 + b_1 , \tag{6}$$

$$\overline{C}_3 = b_0 + b_2 , \tag{7}$$

$$\overline{C}_4 = b_0 . \tag{8}$$

The solution to this set of equations for the two-item case is

$$b_0 = \overline{C}_4 , \tag{9}$$

$$b_1 = \overline{C}_2 - \overline{C}_4 , \tag{10}$$

$$b_2 = \overline{C}_3 - \overline{C}_4 , \tag{11}$$

$$b_{12} = \overline{C}_1 + \overline{C}_4 - \overline{C}_2 - C_3 . \tag{12}$$

In a similar way, if there were three items there would be $2^3 = 8$ unknown parameters in the polynomial prediction equation

$$\hat{C} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3 . \tag{13}$$

Since there are 8 criterion means, this would again lead to the exact solution of a set of 8 equations.

To prove Theorem 2, it is only necessary to show that there will always be at most 2^t unknowns in the polynomial prediction formula. The proof is as follows: there will always be $(t + 1)$ unknowns, since the first term in the polynomial is the unknown constant b_0 and each item score will have an unknown coefficient $b_1 , b_2 , b_3 , \cdots , b_t$. The square, or higher power, of any item score reduces to the item score itself ($0^k = 0$, $1^k = 1$), so all powers of X are simply X, i.e., $X^k = X$. Therefore, powers of item scores need not be considered. There will be $t(t - 1)/2$ cross-product terms of the type X_1X_2 , X_1X_3 , etc., since this is the number of times two objects can be selected from a set of t objects. There will be $t(t - 1)(t - 2)/3!$ cross-products of the type $X_1X_2X_3 , X_1X_2X_4 , X_2X_3X_4$, etc. Since there is one unknown constant for each cross-product term, the total number of unknown coefficients

for all terms is

$$(14) \qquad 1 + t + \frac{t(t-1)}{2} + \frac{t(t-1)(t-2)}{3!} + \cdots = \sum_{i=0}^{t} \frac{t!}{i!(t-i)!} = 2^t .$$

COROLLARY to THEOREM 2. *Whenever the number of empty answer patterns for t items is g, then the number of terms in the polynomial equation will be $2^t - g$.*

Applying this corollary, if the number of filled answer patterns is $(t + 1)$ or less, and the t items are linearly independent, then the linear multiple regression on the t items will give maximum validity. Of course there may be cases where the number of filled answer patterns is more than $(t + 1)$ and the linear regression still gives maximum validity. This case will be discussed in Section 4 in relation to tests of significance.

For example, consider the perfect Guttman scale, which has both properties—there are only $(t + 1)$ answer patterns, and the t items are linearly independent. In Table 1 if the *No-Yes* answer pattern were empty, the two items would form a perfect Guttman scale. The polynomial predictor would become the linear equation

$$(15) \qquad \hat{C} = b_0 + b_1 X_1 + b_2 X_2 ,$$

where

$$(16) \qquad b_0 = \overline{C}_4 ,$$

$$(17) \qquad b_1 = \overline{C}_2 - \overline{C}_4 ,$$

and

$$(18) \qquad b_2 = \overline{C}_1 - \overline{C}_2 .$$

These solutions for the b's make $\hat{C}_i = \overline{C}_i$ for the ith answer pattern.

The Guttman scale score is defined as the sum of the item scores, $S = X_1 + X_2$. In other words, Guttman sets $b_0 = 0$, $b_1 = b_2 = 1$, in (15). Guttman [3, p. 89] states that for a perfect scale, this sum score is sufficient for maximum validity. A multiple regression on the item scores is not necessary because the scale score contains all the necessary information: "The predictability of any outside variable from the scale scores is the same as the predictability from the multivariate distribution with the attributes. The zero-order correlation with the scale score is equivalent to the multiple correlation with the universe. Hence, scale scores provide an invariant quantification of the attributes for predicting any outside variable whatsoever."

This statement is correct only if the phrase "correlation ratio" is substituted for "zero-order correlation." There are many cases where the zero-order correlation of the scale score with the criterion is less than the multiple

correlation of the criterion with the item scores. In order to demonstrate this, examine the perfect Guttman scale shown in Table 2.

TABLE 2
A Perfect Guttman Scale for Two Items

Answer Pattern	Items 1	Items 2	Criterion Average
1	yes	yes	$\bar{C}_1$
2	yes	no	$\bar{C}_2$
3	no	no	$\bar{C}_3$

The best fitting multiple linear regression is

$$\hat{C} = b_0 + b_1X_1 + b_2X_2 , \tag{19}$$

where

$$b_0 = \overline{C}_3 , \tag{20}$$

$$b_1 = \overline{C}_2 - \overline{C}_3 , \tag{21}$$

$$b_2 = \overline{C}_1 - \overline{C}_2 . \tag{22}$$

In order for the scale score to have a validity equal to the multiple correlation, b_1 must equal b_2 . This imposes the restriction that

$$\overline{C}_2 - \overline{C}_3 = \overline{C}_1 - \overline{C}_2 . \tag{23}$$

This is a stringent requirement that cannot always be met. For example, if $\overline{C}_2$ is less than $\overline{C}_1$ and is also less than $\overline{C}_3$, this makes b_1 positive and b_2 negative, and the scale score will tend to have near-zero validity.

The Guttman scale score is a special case of the ordinary total score, where the item scores are simply added together. In (3) for the general configural scale, if $b_1 = b_2 = b_3 = \cdots = b_t = 1$, and all other coefficients equal zero then

$$\hat{C} = X_1 + X_2 + \cdots + X_t . \tag{24}$$

Similarly, the multiple regression scale is the case where only the linear portion of (3) is used, i.e.,

$$\hat{C} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_tX_t . \tag{25}$$

Obviously, the validity of these scales can be ranked as follows:

$$\text{configural} \geq \text{multiple regression} \geq \text{total score.}$$

It should be emphasized, of course, that this statement holds only for the sample being analyzed. Because of the loss of the degrees of freedom for the configural scale, this relation may not be found when the scale derived in one sample is applied to another sample.

4. *Analysis of Variance Tests of Significance*

As has been mentioned previously, the least squares configural scale is a maximum likelihood solution when the distribution of criterion scores within each answer pattern is normal. If further, the 2^t criterion score variances are homogeneous, then the analysis of variance technique can be used for tests of significance. The polynomial function used in this paper can be shown to be an exact algebraic transformation of Fisher's analysis of variance mathematical model for the case of equal answer pattern frequencies. This is true whenever the systematic portion of the analysis of variance model is equal to the cell mean.

These tests can be used to answer such questions as: Is the validity of the configural scale significantly greater than zero? Is the validity of the total score significantly greater than that of the configural scale? Will the linear multiple regression give maximum validity, or are non-linear terms necessary? If m items are added to the test, will the configural scale validity increase? Are there certain terms in the polynomial predictor which do not contribute significantly to the validity? All these questions and other similar ones can be answered by the general F-ratio test.

For example, suppose the question arises, "Is the validity of the configural scale greater than zero?" The exact F-ratio test is

$$(26)^* \qquad F = \left(\frac{B}{2^t - 1}\right)\left(\frac{N - 2^t}{W}\right) = \left(\frac{r^2}{2^t - 1}\right)\left(\frac{N - 2^t}{1 - r^2}\right)$$

with $(2^t - 1)$ over $(N - 2^t)$ degrees of freedom.

Some definitions are needed to make the terms in (26) clear from a computational point of view. Let

t be the number of dichotomous items,
N be the total number of individuals
n_i be the number of subjects in the ith answer pattern,
C_{ij} be the criterion score for the jth individual in the ith answer pattern,
$\overline{C}_{i.}$ be the average criterion score for the ith answer pattern, i.e., the configural score for the ith answer pattern,
$\overline{C}_{..}$ be the average criterion score for all N individuals.

*In formula (26) and elsewhere it is assumed that all 2^t answer patterns are filled. If g answer patterns are empty, then the degrees of freedom for W equals $(N - 2^t + g)$ and the degrees of freedom for B equals $(2^t - 1 - g)$.

Then B is the *between* sum of squares, i.e.,

$$B = \sum_{i=1}^{2^t} n_i(\bar{C}_{i.} - \bar{C}_{..})^2. \tag{27}$$

W is the *within* or residual sum of squares, i.e.,

$$W = \sum_{i=1}^{2^t} \sum_{j=1}^{n_i} (C_{ij} - \bar{C}_{i.})^2, \tag{28}$$

and

$$r^2 = B/T = \eta^2. \tag{29}$$

Similarly one can find out if S, the total score obtained by just adding up the unweighted item scores, is sufficient to produce maximum validity. Let r_s be the correlation of S with the criterion; then

$$F = \left(\frac{\eta^2 - r_s^2}{2^t - 2}\right)\left(\frac{N - 2^t}{1 - \eta^2}\right) \tag{30}$$

with $(2^t - 2)$ over $(N - 2^t)$ degrees of freedom.

If the configural scale validity, η, is significantly greater than that of the total score, the next question which can be raised is whether the relation between the item responses and the criterion is linear or non-linear. It is suspected that with most current mental tests, there is a linear relation. This question can be answered by seeing if η^2, the squared configural scale validity, is significantly larger than R^2, the squared multiple correlation based on the t items.

$$F = \left(\frac{\eta^2 - R^2}{2^t - t - 1}\right)\left(\frac{N - 2^t}{1 - \eta^2}\right) \tag{31}$$

with $(2^t - t - 1)$ over $(N - 2^t)$ degrees of freedom.

Another question that can be answered is as follows: suppose m items are added to a k-item test. Is the validity of the $(k + m)$-item configural scale greater than that of the k-item configural scale? The F-ratio significance test is

$$F = \left(\frac{W_k - W_{k+m}}{2^{k+m} - 2^k}\right)\left(\frac{N - 2^{k+m}}{W_{k+m}}\right) = \left(\frac{r_{k+m}^2 - r_k^2}{2^{k+m} - 2^k}\right)\left(\frac{N - 2^{k+m}}{1 - r_{k+m}^2}\right), \tag{32}$$

where W_k and r_k^2 refer to statistics calculated on the basis of the k-item configural scale, W_{k+m} and r_{k+m}^2 refer to statistics calculated on the basis of the $(k + m)$-item configural scale, and the degrees of freedom are $(2^{k+m} - 2^k)$ over $(N - 2^{k+m})$.

In general, it is possible to test the significance of any particular subset of terms in the polynomial predictor. Let H_0 refer to any (null) hypothesis which restricts some of the parameters of the polynomial predictors on an

a priori basis. Let H_1 refer to the non-null hypothesis which places no restrictions on the 2^t parameters. Then the general formula for F is

$$F = \left(\frac{W_0 - W_1}{v_0 - v_1}\right)\left(\frac{v_1}{W_1}\right), \tag{33}$$

where $v_0 \equiv N$ minus the number of parameters used in predicting the criterion scores according to the H_0 ,

$W_0 \equiv$ the deviance (sum of squared errors of prediction) obtained by applying H_0 ,

$v_1 \equiv N - 2^t$,

$W_1 \equiv$ the deviance obtained by applying the polynomial predictor, and the degrees of freedom are $(v_0 - v_1)$ over $(N - v_1)$.

Another way of writing it is

$$F = \left(\frac{\eta_1^2 - \eta_0^2}{v_0 - v_1}\right)\left(\frac{v_1}{1 - \eta_1^2}\right), \tag{34}$$

where

$$\eta_1^2 = \frac{T - W_1}{T},$$

$$\eta_0^2 = \frac{T - W_0}{T},$$

and T is the deviance about the general mean.

Equations (33) and (34) give the general solution for testing what are known as "linear hypotheses," [5, pp. 298-302]. This allows the reader to construct his own test of significance for any question about the polynomial predictors.

5. *Discussion*

Many ingenious methods of pattern and profile analysis are being used today in an attempt to increase the predictability of the criterion. Gaier and Lee [1] in a partial review of the literature, summarized some 28 references. Presumably, one could take a set of data and compare all of the known methods to see which has the greatest validity. This would be a laborious and inefficient way of solving the problem. As Horst has said [4, p. 8], "The work in this area will be much more fruitful when more precise and rigorous mathematical concepts are developed to take the place of verbal formulations and analyses based on empirical or trial and error manipulations of the data."

Given the case of t dichotomous items and a quantitative criterion, the least squares approach shows that the configural scale, a tth degree polynomial function of the t item scores, possesses maximum validity. To the extent that any of the present techniques of pattern analysis can reach this maximum validity, they are special cases of the polynomial function.

It is always possible to write a mathematical description of the configural scale and apply the usual matrix algebra theorems to the result. For example, the polynomial predictor in matrix form is

$$\overline{C} = Xb, \tag{35}$$

where $\overline{C}$ is the 2^t by one column vector of observed criterion averages,
X is a 2^t by 2^t matrix of zero-one entries, where the rows represent answer patterns and the columns represent terms of the polynomial,
b is the 2^t by one column of coefficients.

If all the 2^t answer patterns are filled, and X is non-singular, then

$$b = (X^{-1})\overline{C} \tag{36}$$

is an exact solution. If some of the answer patterns are empty, X can still be made into a square non-singular matrix by eliminating the corresponding polynomial terms and the least squares solution is still as above.

The principal advantage of calculating b is that the relative importance of each term in the polynomial predictor is specified numerically. For example, suppose the criterion is not a function of the second-order interactions of the items. Then, when b is calculated, all coefficients of the type b_{ij} will be near zero. In such a case, when it has been hypothesized that only k of the regression coefficients are non-zero, a least squares solution can be obtained which uses only the k specified terms.

It is convenient to compute first an approximate least squares solution. If this solution gives an adequate fit, then only the k specified terms are needed. If the approximate solution does not give an adequate fit, then it is necessary to compute the exact least squares solution.

The approximate least squares solution is as follows: Let X_k be the 2^t by k matrix obtained by selecting the specified k columns from X. Then

$$b_k = (X_k'X_k)^{-1}X_k'\overline{C}. \tag{37}$$

The exact least squares solution is as follows: Let Z_k be an N by k matrix whose general element, x_{jm} , is the score of the jth individual on the mth term of the polynomial. Then, given that the jth individual is in the ith answer pattern, the jth row of Z_k is exactly equal to the ith row of X_k . Essentially Z_k is an expanded form of X_k where each of the rows of X_k has been repeated n_i times. Let C be an N-rowed column vector where C_j is the criterion score of the jth individual. Then

$$w_k = (Z_k'Z_k)^{-1}Z_k'C \tag{38}$$

is the set of regression coefficients which give the exact least squares fit.

Equation (38) provides a test on whether any specified set of item interactions is related to the criterion. It can be a powerful tool for testing psychological hypotheses about the relation between subject's responses to

the items and his criterion score. This could be the most useful function of configural analysis.

Another possibility is to use the configural technique for item analysis. This would involve an empirical search pattern with or without the use of hypotheses. However, considerable caution is needed in such empirical applications of configural scoring since the degrees of freedom used for computing the regression coefficients increase exponentially with the number of selected items. In general, any procedure involving configural scoring implies a very small number of items and a large number of subjects. Also, unless the items have been specially constructed, or there are good theoretical grounds for believing that non-linear relations exist, the usual total score will probably give maximum validity.

It is especially necessary to be careful in generalizing from the analysis sample to future samples. Because of loss of degrees of freedom, there will be a sizeable decrease in the cross-validity.

One procedure for guarding against errors in generalization to future samples would be as follows: (a) Compute back-validity (validity on the analysis sample) and test whether it differs from zero. (b) If the back-validity is significantly greater than zero, test to see if it is significantly greater than the multiple correlation and total score back-validities. (c) If the above tests are positive, the cross-validities for each model should be estimated by Lord's method [6] to see if the configural scale has any practical advantage. (d) As a final safeguard, the actual cross-validities can be computed and tested for significance. If the estimated cross-validity differences fall to zero, then it is unnecessary to analyze the cross-validation sample.

The authors are grateful to Professor Charles F. Wrigley and Dr. Maurice Lorr for advice and critical comment; and to Mrs. Ruth Heitman for her typing services.

REFERENCES

[1] Gaier, E. L. and Lee, Marilyn C. Pattern analysis. The configural approach to predictive measurement. *Psychol. Bull.* 1953, **50**, 141-149.

[2] Guttman, L. In The prediction of personal adjustment, P. Horst (Ed.), Bulletin 48, Social Science Research Council, New York, 1941.

[3] Guttman, L. In Measurement and prediction, S. A. Stouffer, F. A. Suchman, P. F. Lazarsfeld, *et al.* Princeton, New Jersey: Princeton University Press, 1950.

[4] Horst, P. Pattern analysis and configural scoring. *J. clin. Psychol.*, 1954, **10**, 3-11.

[5] Kendall, M. G. The advanced theory of statistics. Vol. II. London: Griffin, 1946.

[6] Lord, F. M. Efficiency of prediction when a regression equation from one sample is used in a new sample. RB 50-40, Educational Testing Service, 1950.

[7] Meehl, P. E. Configural scoring. *J. consult. Psychol.*, 1950, **14**, 165-171.

[8] Rao, C. R. Utilization of multiple measurements in problems of biological classification. *J. roy. statist. Soc. B*, 1948, **10**, 159-203.

Manuscript received 1/9/56
Revised manuscript received 3/5/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

THE EXPECTED VARIANCE OF THE SAMPLING ERRORS FOR A SET OF ITEM-CRITERION CORRELATIONS

HUBERT E. BROGDEN

PERSONNEL RESEARCH BRANCH

THE ADJUTANT GENERAL'S OFFICE*

An expression for the expected variance of the sampling errors for the validities of a set of correlated items that is computationally feasible when the number of items is large is developed. Since the item difficulties are assumed to be constant, the estimate must be applied to pools or sub-pools of items reasonably homogeneous with respect to difficulty.

In item analysis and in a number of related problems, an expression giving the variance of the sampling errors of a set of item validities would often be useful. The standard error of the item validity coefficients is not a satisfactory estimate since it is well known that the sampling errors in the validities of a set of correlated variables are themselves correlated. Wishart [1] has presented a general though complex solution giving the sampling distribution of the covariance matrix for a set of correlated variables. His solution is not, however, feasible and may not be appropriate when applied to the problems arising in dealing with the sampling variation of a set of item validities. This note will propose a solution to the problem of estimating the variance, but not the distribution, of the sampling errors for a set of item validities that appears to be both simple and feasible.

At least one application of such an estimate is obvious. Many investigators, upon finding that 5 per cent of a set of items are valid at the 5 per cent level of confidence, conclude that their item analysis data are of no value. It is hoped that the sampling estimate to be presented will permit a more accurate conclusion in problems of this nature.

Definition of Symbols

$r_i \equiv r_{x_i y}$, the point-biserial (or product-moment) correlation, in the sample, between any of a set of items and an external criterion. The items $(x_1 , x_2 , \cdots , x_i , \cdots , x_n)$ take on values of 1 for the correct choice and 0 for the incorrect choice. The correct alternative must be determined, though arbitrarily, before computing the item validities.

*The opinions expresssd are those of the author and do not reflect official Department of the Army policy.

$\tilde{r}_i \equiv \tilde{r}_{x_i y}$, the point-biserial correlation between any item and the criterion in the universe.

$\sigma_i \equiv \sigma_{x_i}, \sqrt{P_i Q_i}$, the standard deviation of an item score in a sample.

$\tilde{\sigma}_i \equiv$ the standard deviation of an item score in the universe.

$t = \sum_i x_i$, the sum of the item scores—a score obtained by summing the number of correct (as defined above) alternatives.

$N \equiv$ the number of individuals in the sample.

$E(\sigma^2_{(r_i - \tilde{r}_i)}) \equiv$ the expected variance of the errors in a set of item validities for a sample.

Assumptions

1. It is assumed that all items have equal difficulty and, as a corollary, that σ_i is constant across items.

2. It is assumed that σ_i and σ_t are satisfactory estimates of, respectively, the $\tilde{\sigma}_i$ and $\tilde{\sigma}_t$. This assumption is similar to, but less restrictive than, the usual assumption in similar developments that the predictors remain fixed in going from the sample to the universe.

3. It is assumed that t and the criterion are normally distributed.

The Derivation

The problem is to determine, in a form feasible for calculation, an expression for the expected variance of the errors for a set of item validities in a sample. Since the errors are the discrepancies between the universe and sample values, the expected value of the variance of the errors in the validities of a set of items has the basic definition

$$(1) \qquad E(\sigma^2_{(r_i - \tilde{r}_i)}) = E\{(1/n) \sum_i (r_i - \tilde{r}_i)^2 - [(1/n) \sum_i (r_i - \tilde{r}_i)]^2\}$$

$$(2) \qquad = E[(1/n) \sum_i (r_i - \tilde{r}_i)^2] - E[(1/n) \sum_i r_i - (1/n) \sum_i \tilde{r}_i]^2 .$$

From a well-known formula, the correlation of the sum of the items (t) with the criterion may be written, if σ_i is assumed to be constant,

$$(3) \qquad r_{ty} = \sigma_i \sum_i r_i / \sigma_t ,$$

and, consequently,

$$(4) \qquad r_{ty}(\sigma_i / \sigma_i) = \sum_i r_i .$$

By similar reasoning

$$(5) \qquad \tilde{r}_{ty}(\tilde{\sigma}_i / \tilde{\sigma}_i) = \sum_i \tilde{r}_i .$$

Substituting in (2)

$$E(\sigma^2_{(r_i - \tilde{r}_i)}) = E[(1/n) \sum_i (r_i - \tilde{r}_i)^2]$$

(6)

$$- E[(1/n) r_{tv} (\sigma_t / \sigma_i) - (1/n) \tilde{r}_{tv} (\tilde{\sigma}_t / \tilde{\sigma}_i)]^2.$$

From assumption 2, $\tilde{\sigma}_t$ equals σ_t and $\tilde{\sigma}_i$ equals σ_i . Reducing further,

$$(7) \qquad E(\sigma^2_{(r_i - \tilde{r}_i)}) = (1/n) \sum_i E(r_i - \tilde{r}_i)^2 - (1/n^2)(\sigma_t^2 / \sigma_i^2) E(r_{tv} - \tilde{r}_{tv})^2.$$

The expected value of the square of $(r_i - \tilde{r}_i)$ or of $(r_{tv} - \tilde{r}_{tv})$ is the average of their squared values over an infinite series of samples and is, consequently, also equal to their sampling variance. Formulas are available to evaluate them, and σ_t can be obtained from the sample. Thus, (7) provides a solution to the original problem, although this solution may be somewhat tedious since the sampling variance for each item validity must be determined. In most item analysis problems, the r_i should be sufficiently close to zero so that $1/N$ may be regarded as a satisfactory estimate of their sampling variance. In this event, (7) will reduce to

$$(8) \qquad E(\sigma^2_{(r_i - \tilde{r}_i)}) = (1/N)[1 - (1 - \tilde{r}^2_{tv})(\sigma_t^2 / n^2 \sigma_i^2)].$$

(8) gives a feasible solution to the original problem.

(8) will reduce further if the *average* item intercorrelation is zero. If this is true, the average of the off-diagonal entries of the full symmetric matrix of item covariances (whose sum equals σ_t^2) will also equal zero, and σ_t^2 will reduce to $n\sigma_i^2$. If n is large,

$$(9) \qquad E(\sigma^2_{(r_i - \tilde{r}_i)}) = 1/N$$

(9) is, of course, the expected result if full statistical independence of the items is assumed. The derivation just presented shows that a less restrictive assumption permits the use of this simple formula. In practice, if σ_t^2 does not exceed $n\sigma_i^2$, $1/N$ may be used in place of (8) to give an estimate of the variance of the sampling errors across a set of items. It should be stressed that the less restrictive assumption applies in estimating the variance of the sampling errors; nothing has been demonstrated regarding the distribution of sampling errors.

While the writer had primary interest in the expected variance of the sampling errors of the validities of a set of dichotomous items, an adaptation of (7) will apply to the validities of a set of continuous predictors. The assumption of equal item difficulty is, of course, unnecessary. It must be assumed that the standard deviations of the individual predictors and the standard deviation of the sum of the predictors are the same in the sample and universe and that t is redefined as the sum of the predictors with each predictor

converted to unit standard deviation form. With these assumptions,

$$E(\sigma^2_{(r_i - \tilde{r}_i)}) = (1/n) \sum_i E(r_i - \tilde{r}_i)^2 - \sigma_i^2(1 - r^2_{i_y})/n^2N. \tag{10}$$

Discussion

To apply the formula given in (8), t must be determined for each case in the sample, and σ_i must be computed. The scoring run to compute t is the greater part of the labor involved. By the definition given to t, the sign of an item in such a scoring run must correspond to the sign used when the item validities were initially computed.

The formula behaves as would be expected in those special cases where the solution is obvious. If the intercorrelations of the items are all plus one, the variance of the errors of a set of item validities should be zero and the solution by the formula yields zero variance of the errors. If the items are independent of each other, the errors should be independent of each other, as shown in (9), and the formula should and does simplify to the sampling variance of a correlation coefficient.

Since the p-values of the items were assumed to be constant in deriving the formula, a pool of items involving a considerable range of difficulty will have to be subdivided into pools of constant item difficulty before the formula is applied. In the author's opinion, a range of difficulty of at least .10 can be permitted without introducing serious error, since the variation of σ_i is quite small within such a p-value range.

REFERENCE

[1] Wishart, J. The generalized product-moment distribution in samples from a normal multivariate population. *Biometrika*, 1928, **20A**, 32.

Manuscript received 5/6/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

A NECESSARY AND SUFFICIENT FORMULA FOR MATRIC FACTORING

LOUIS GUTTMAN

CENTER FOR ADVANCED STUDY IN THE BEHAVIORAL SCIENCES*

For the purpose of extracting factors from matrices, it is proved that a certain formula is both necessary and sufficient. In factor analysis, the formula may be applied either to the correlation matrix, or directly to the score matrix (assuming the communality problem is solved). As many factors as desired can be extracted in one operation. Having such a compact formulation is useful for teaching as well as computing purposes, since it includes all techniques of factor extraction as special cases.

Let A be an arbitrary (real) matrix of order $p \times q$ and rank r. It is desired to extract factors from A by finding a matrix A_1 of rank s, where $s \leqq r$, of the form

$$A_1 = BDC, \tag{1}$$

where D is non-singular and of order s, such that A_2 shall be of rank $r - s$, where

$$A_2 = A - A_1 . \tag{2}$$

This requires further that B and C be of rank s and of orders $p \times s$ and $s \times q$, respectively.

Such a problem occurs in factor analysis in at least two different but closely related ways:

(a) A may be the observed score matrix after unique-factor scores are subtracted out, for q individuals on p tests. Then B (or BD) can be regarded as common-factor loadings of the tests, and DC (or C) as common-factor scores of the respondents.

(b) A may be the observed correlation matrix with communalities in the main diagonal. In this case, $p = q$; $C = B'$; and A, A_1 , A_2 , and D are restricted to being Gramian. Then B again gives common-factor loadings of the tests, while now D is the inverse of the covariance matrix of the common factors, being a diagonal matrix when the common factors are orthogonal.

*On leave from the Israel Institute for Applied Social Research. This research was facilitated in part by a grant from the Lucius N. Littauer Foundation to the American Committee for Social Research in Israel, Inc.

In either case, when $s = 1$, a single factor is extracted from A by (2), as by the centroid method, principal axis, or other ways of reducing r by 1. When $s > 1$, several factors are extracted simultaneously, as in multiple-group methods. When $s = r$, or $A_2 = 0$, all factors are extracted in one step (cf. [1, 2]). The relationship between the factoring of scores and the factoring of correlation coefficients has been analyzed in [1, 2].

It has been shown in [1] that a *sufficient* formula for A_1 is as follows. Let X and Y be arbitrary weight matrices of orders $s \times p$ and $s \times q$, respectively, and such that XAY' is non-singular. Let

$$D = (XAY')^{-1}. \tag{3}$$

Thus, D is of rank and order s. Compute B and C by the formulas

$$B = AY', \qquad C = XA. \tag{4}$$

Then if A_1 is computed by formula (1), A_1 must be of rank s and A_2 in (2) of rank $r - s$. For factoring the correlation matrix as in case (b) above, let $Y = X$.

This sufficient technique for extracting factors is actually only a generalization of Lagrange's technique for reducing bilinear forms, as pointed out in [1].

It is of considerable interest to inquire* as to whether any other kind of formula is possible for A_1 in (1), keeping D non-singular, but removing conditions (3) and (4). An important restriction in (4) is that the factor matrices B and C are linear transformations of A. Is it possible for factors to exist that are not such functions of A?

The answer turns out to be in the negative. If A_1 is of rank s and reduces the rank of A to $r - s$, then A_1 in the form (1) must *always* have B, C, and D of the forms (4) and (3). Our formulas are *necessary* as well as sufficient.

For the proof, suppose A_1 is of the form (1) and is of rank s, and D is non-singular of order s. Thus, B and C are of orders $p \times s$ and $s \times q$, respectively. Define the partitioned matrix E to be

$$E = \begin{bmatrix} A & B \\ C & D^{-1} \end{bmatrix}. \tag{5}$$

E is A enlarged by s rows and s columns. By direct multiplication it is verified that

$$E = \begin{bmatrix} I_p & B \\ 0 & D^{-1} \end{bmatrix} \cdot \begin{bmatrix} A_2 & 0 \\ DC & I_s \end{bmatrix}, \tag{6}$$

where I_p and I_s are the unit matrices of order p and s, respectively, and A_2 is the residual matrix defined by (2).

*This problem was suggested to the writer by Dr. W. A. Gibson.

Let t be the rank of A_2. Since the first matrix on the right of (6) is clearly non-singular, and the rank of the second matrix is clearly $s + t$ (the sum of the ranks of A_2 and I_s), the rank of E must be $s + t$. Therefore, a necessary and sufficient condition that $t = r - s$ is that the rank of E equal r. But in the right of (5), A by itself is already of rank r, so a necessary and sufficient condition that E be of rank r is that the last submatric row in the right of (5) be linearly dependent on the first submatric row, or that there exist an X such that

$$C = XA, \quad D^{-1} = XB. \tag{7}$$

Similarly, the last submatric column must be linearly dependent on the first submatric column, or there exists a Y such that

$$B = AY', \quad D^{-1} = CY'. \tag{8}$$

Note that X and Y need not be uniquely determined when $p > r$ and $q > r$, respectively. The first parts of (7) and (8) yield (4); substituting the first part of (7) in the last part of (8) yields (3).

Thus, simultaneously both the necessity and sufficiency of the factoring formulas, in place of only the sufficiency proof in [1], have been proved. All possible factoring methods, whether directly on the score matrix or on the correlation matrix, can differ only in the choice of weight matrices X and Y. This fact not only gives a unified and simplified approach to practical computing procedures (cf. [2, 4]), but also—as Lubin has pointed out—serves as a simple basis for teaching factor analysis to beginning students [5].

It must be cautioned, however, that the above formulas assume the communality problem solved. The gravity of this assumption is analyzed in [3].

REFERENCES

[1] Guttman, L. General theory and methods for matric factoring. *Psychometrika*, 1944, **9**, 1-16.

[2] Guttman, L. Multiple group methods for common-factor analysis: their basis, computation, and interpretation. *Psychometrika*, 1952, **17**, 209-222.

[3] Guttman, L. The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *Brit. J. statist. Psychol.*, 1955, **8**, 65-81.

[4] Harman, H. H. The square root method and multiple group methods of factor analysis. *Psychometrika*, 1954, **19**, 39-55.

[5] Lubin, A. Review of Fruchter's Introduction to Factor Analysis, *Personnel Psychology*, 1954, **7**, 577-581.

Manuscript received 9/19/55

Revised manuscript received 12/14/55

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

EXACT PROBABILITIES FOR CONTINGENCY TABLES USING BINOMIAL COEFFICIENTS

JAMES M. SAKODA
AND
BURTON H. COHEN

UNIVERSITY OF CONNECTICUT

The use of binomial coefficients in place of factorials to shorten the calculation of exact probabilities for 2×2 and $2 \times r$ contingency tables is discussed. A useful set of inequalities for estimating the cumulative probabilities in the tail of the distribution from the probability of a single table is given. A table of binomial coefficients with four significant places and n through 60 is provided.

A 2×2 contingency table and a numerical example are represented as follows:

a	b	$a+b$	7	8	15
c	d	$c+d$	8	37	45
$a+c$	$b+d$	N	15	45	60

Under the hypothesis of independence the exact probability, p, of specified values of a, b, c, d given the marginal totals $(a+b)$, $(c+d)$, $(a+c)$, $(b+d)$ can be written either in terms of factorials or binomial coefficients [1, 2]:

$$p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\;\;b!\;\;c!\;\;d!\;\;N!} = \frac{{}_{a+b}C_a\;{}_{c+d}C_c}{{}_NC_{a+c}}. \tag{1}$$

Using binomial coefficients for our example,

$$p = \frac{6435 \cdot 2156 \cdot 10^5}{5319 \cdot 10^{10}} = .02608.$$

Use of binomial coefficients in the calculation of cumulative probabilities, P, for a given table and those more extreme than it permits the possibility of cumulating cross products on a desk calculator as follows:

$$P = \frac{\sum {}_{a+b}C_a\;{}_{c+d}C_c}{{}_NC_{a+c}}. \tag{2}$$

For the numerical example

$$P = \frac{{}_{15}C_7\;{}_{45}C_8 + {}_{15}C_8\;{}_{45}C_7 + {}_{15}C_9\;{}_{45}C_6 + \cdots}{{}_{60}C_{15}}$$

$$= \frac{6435 \cdot 2156 \cdot 10^5 + 6435 \cdot 4538 \cdot 10^4 + 5005 \cdot 8145 \cdot 10^3}{5319 \cdot 10^{10}}$$

$$= .03234.$$

This is the probability for a one-tailed test based on three terms only. No more than four or five terms are generally necessary to obtain a fairly accurate probability. It can be shown that in the critical region of the distribution the

Binomial Coefficients, ${}_nC_r$

r or n - r	n = 1	n = 2	n = 3	n = 4	n = 5	n = 6	n = 7	n = 8	n = 9	n = 10
1	1	2	3	4	5	6	7	8	9	10
2		1	3	6	10	15	21	28	36	45
3			1	4	10	20	35	56	84	120
4				1	5	15	35	70	126	210
5					1	6	21	56	126	252

r or n - r	n = 11	n = 12	n = 13	n = 14	n = 15	n = 16	n = 17	n = 18	n = 19	n = 20
1	11	12	13	14	15	16	17	18	19	20
2	55	66	78	91	105	120	136	153	171	190
3	165	220	286	364	455	560	680	816	969	1140
4	330	495	715	1001	1365	1820	2380	3060	3876	4845
5	462	792	1287	2002	3003	4368	6188	8568	1163-1	1550-1
6	462	924	1716	3003	5005	8008	1238-1	1856-1	2713-1	3876-1
7	330	792	1716	3432	6435	1144-1	1945-1	3182-1	5039-1	7752-1
8	165	495	1287	3003	6435	1287-1	2431-1	4376-1	7558-1	1260-2
9	55	220	715	2002	5005	1144-1	2431-1	4862-1	9238-1	1680-2
10	11	66	286	1001	3003	8008	1945-1	4376-1	9238-1	1848-2

r or n - r	n = 21	n = 22	n = 23	n = 24	n = 25	n = 26	n = 27	n = 28	n = 29	n = 30
1	21	22	23	24	25	26	27	28	29	30
2	210	231	253	276	300	325	351	378	406	435
3	1330	1540	1771	2024	2300	2600	2925	3276	3654	4060
4	5985	7315	8855	1063-1	1265-1	1495-1	1755-1	2048-1	2375-1	2741-1
5	2035-1	2633-1	3365-1	4250-1	5313-1	6578-1	8073-1	9828-1	1188-2	1425-2
6	5426-1	7461-1	1009-2	1346-2	1771-2	2302-2	2960-2	3767-2	4750-2	5938-2
7	1163-2	1705-2	2452-2	3461-2	4807-2	6578-2	8880-2	1184-3	1561-3	2036-3
8	2035-2	3198-2	4903-2	7355-2	1082-3	1562-3	2220-3	3108-3	4292-3	5853-3
9	2939-2	4974-2	8172-2	1308-3	2043-3	3125-3	4687-3	6907-3	1002-4	1431-4
10	3527-2	6466-2	1144-3	1961-3	3269-3	5312-3	8436-3	1312-4	2003-4	3005-4
11	3527-2	7054-2	1352-3	2496-3	4457-3	7726-3	1304-4	2147-4	3460-4	5463-4
12	2939-2	6466-2	1352-3	2704-3	5200-3	9658-3	1738-4	3042-4	5190-4	8649-4
13	2035-2	4974-2	1144-3	2496-3	5200-3	1040-4	2006-4	3744-4	6786-4	1198-5
14	1163-2	3198-2	8172-2	1961-3	4457-3	9658-3	2006-4	4012-4	7756-4	1454-5
15	5426-1	1705-2	4903-2	1308-3	3269-3	7726-3	1738-4	3744-4	7756-4	1551-5

r or n - r	n = 31	n = 32	n = 33	n = 34	n = 35	n = 36	n = 37	n = 38	n = 39	n = 40
1	31	32	33	34	35	36	37	38	39	40
2	465	496	528	561	595	630	666	703	741	780
3	4495	4960	5456	5984	6545	7140	7770	8436	9139	9880
4	3147-1	3596-1	4092-1	4638-1	5236-1	5890-1	6605-1	7382-1	8225-1	9139-1
5	1699-2	2014-2	2373-2	2783-2	3246-2	3770-2	4359-2	5019-2	5758-2	6580-2
6	7363-2	9062-2	1108-3	1345-3	1623-3	1948-3	2325-3	2761-3	3263-3	3838-3
7	2630-3	3366-3	4272-3	5380-3	6725-3	8348-3	1030-4	1262-4	1538-4	1864-4
8	7889-3	1052-4	1388-4	1816-4	2354-4	3026-4	3861-4	4890-4	6152-4	7690-4
9	2016-4	2805-4	3857-4	5245-4	7061-4	9414-4	1244-5	1630-5	2119-5	2734-5
10	4435-4	6451-4	9256-4	1311-5	1836-5	2542-5	3483-5	4727-5	6357-5	8477-5
11	8467-4	1290-5	1935-5	2861-5	4172-5	6008-5	8550-5	1203-6	1676-6	2312-6
12	1411-5	2258-5	3548-5	5484-5	8345-5	1252-6	1852-6	2707-6	3911-6	5587-6
13	2063-5	3474-5	5732-5	9280-5	1476-6	2311-6	3562-6	5415-6	8122-6	1203-7
14	2652-5	4714-5	8188-5	1392-6	2320-6	3796-6	6107-6	9670-6	1508-7	2321-7
15	3005-5	5657-5	1037-6	1856-6	3248-6	5568-6	9364-6	1547-7	2514-7	4023-7
16	3005-5	6011-5	1167-6	2204-6	4060-6	7308-6	1288-7	2224-7	3771-7	6285-7
17	2652-5	5657-5	1167-6	2334-6	4538-6	8597-6	1591-7	2878-7	5102-7	8873-7
18	2063-5	4714-5	1037-6	2204-6	4538-6	9075-6	1767-7	3358-7	6236-7	1134-8
19	1411-5	3474-5	8188-5	1856-6	4060-6	8597-6	1767-7	3535-7	6892-7	1313-8
20	8467-4	2258-5	5732-5	1392-6	3248-6	7308-6	1591-7	3358-7	6892-7	1378-8

error in omitting terms will be smaller than the probability of the last table which is utilized and equal to or larger than the probability of the first table which is not utilized. If the use of another term produces negligible change in the P value, no further calculation is necessary. In the example, the

Binomial Coefficients, ${}_nC_r$

r or n - r	n = 41	n = 42	n = 43	n = 44	n = 45	n = 46	n = 47	n = 48	n = 49	n = 50
1	41	42	43	44	45	46	47	48	49	50
2	820	861	903	946	990	1035	1081	1128	1176	1225
3	1066-1	1148-1	1234-1	1324-1	1419-1	1518-1	1621-1	1730-1	1842-1	1960-1
4	1013-2	1119-2	1234-2	1358-2	1490-2	1632-2	1784-2	1946-2	2119-2	2303-2
5	7494-2	8507-2	9626-2	1086-3	1222-3	1371-3	1534-3	1712-3	1907-3	2119-3
6	4496-3	5246-3	6096-3	7059-3	8145-3	9367-3	1074-4	1227-4	1398-4	1589-4
7	2248-4	2698-4	3222-4	3832-4	4538-4	5352-4	6289-4	7363-4	8590-4	9988-4
8	9555-4	1180-5	1450-5	1772-5	2156-5	2609-5	3145-5	3773-5	4510-5	5369-5
9	3503-5	4459-5	5639-5	7089-5	8862-5	1102-6	1363-6	1677-6	2054-6	2505-6
10	1121-6	1471-6	1917-6	2481-6	3190-6	4076-6	5178-6	6541-6	8218-6	1027-7
11	3159-6	4281-6	5752-6	7669-6	1015-7	1334-7	1742-7	2260-7	2914-7	3735-7
12	7899-6	1106-7	1534-7	2109-7	2876-7	3891-7	5225-7	6967-7	9226-7	1214-8
13	1762-7	2552-7	3658-7	5192-7	7301-7	1018-8	1407-8	1929-8	2626-8	3549-8
14	3524-7	5286-7	7838-7	1150-8	1669-8	2399-8	3416-8	4823-8	6752-8	9378-8
15	6343-7	9867-7	1515-8	2299-8	3449-8	5117-8	7516-8	1093-9	1576-9	2251-9
16	1031-8	1665-8	2652-8	4167-8	6466-8	9915-8	1503-9	2255-9	3348-9	4924-9
17	1516-8	2547-8	4212-8	6864-8	1103-9	1750-9	2741-9	4244-9	6499-9	9847-9
18	2021-8	3537-8	6084-8	1030-9	1716-9	2819-9	4569-9	7310-9	1155-10	1805-10
19	2447-8	4468-8	8005-8	1409-9	2438-9	4154-9	6973-9	1154-10	1885-10	3041-10
20	2691-8	5138-8	9606-8	1761-9	3170-9	5608-9	9762-9	1674-10	2828-10	4713-10
21	2691-8	5383-8	1052-9	2013-9	3774-9	6944-9	1255-10	2231-10	3905-10	6733-10
22	2447-8	5138-8	1052-9	2104-9	4117-9	7890-9	1483-10	2739-10	4970-10	8875-10
23	2021-8	4468-8	9606-8	2013-9	4117-9	8233-9	1612-10	3096-10	5834-10	1080-11
24	1516-8	3537-8	8005-8	1761-9	3774-9	7890-9	1612-10	3225-10	6321-10	1215-11
25	1031-8	2547-8	6084-8	1409-9	3170-9	6944-9	1483-10	3096-10	6321-10	1264-11

r or n - r	n = 51	n = 52	n = 53	n = 54	n = 55	n = 56	n = 57	n = 58	n = 59	n = 60
1	51	52	53	54	55	56	57	58	59	60
2	1275	1326	1378	1431	1485	1540	1596	1653	1711	1770
3	2083-1	2210-1	2343-1	2480-1	2623-1	2772-1	2926-1	3086-1	3251-1	3422-1
4	2499-2	2707-2	2928-2	3163-2	3411-2	3673-2	3950-2	4243-2	4551-2	4876-2
5	2349-3	2599-3	2870-3	3163-3	3479-3	3820-3	4187-3	4582-3	5006-3	5462-3
6	1801-4	2036-4	2296-4	2583-4	2899-4	3247-4	3629-4	4048-4	4506-4	5006-4
7	1158-5	1338-5	1541-5	1771-5	2029-5	2319-5	2644-5	3007-5	3411-5	3862-5
8	6368-5	7525-5	8863-5	1040-6	1218-6	1420-6	1652-6	1917-6	2217-6	2559-6
9	3042-6	3679-6	4432-6	5318-6	6358-6	7576-6	8996-6	1065-7	1257-7	1478-7
10	1278-7	1582-7	1950-7	2393-7	2925-7	3561-7	4318-7	5218-7	6283-7	7539-7
11	4763-7	6040-7	7622-7	9572-7	1197-8	1489-8	1845-8	2277-8	2799-8	3427-8
12	1588-8	2064-8	2668-8	3430-8	4387-8	5584-8	7073-8	8918-8	1119-9	1399-9
13	4763-8	6350-8	8414-8	1108-9	1451-9	1890-9	2448-9	3156-9	4047-9	5167-9
14	1293-9	1769-9	2404-9	3245-9	4354-9	5805-9	7695-9	1014-10	1330-10	1735-10
15	3189-9	4481-9	6250-9	8654-9	1190-10	1625-10	2206-10	2975-10	3990-10	5319-10
16	7175-9	1036-10	1484-10	2109-10	2975-10	4165-10	5790-10	7996-10	1097-11	1496-11
17	1477-10	2195-10	3231-10	4715-10	6825-10	9800-10	1396-11	1975-11	2775-11	3872-11
18	2790-10	4267-10	6462-10	9693-10	1441-11	2123-11	3103-11	4500-11	6475-11	9250-11
19	4846-10	7636-10	1190-11	1836-11	2806-11	4247-11	6370-11	9473-11	1397-12	2045-12
20	7754-10	1260-11	2024-11	3214-11	5050-11	7856-11	1210-12	1847-12	2795-12	4192-12
21	1145-11	1920-11	3180-11	5203-11	8417-11	1347-12	2132-12	3343-12	5190-12	7984-12
22	1561-11	2705-11	4625-11	7805-11	1301-12	2143-12	3489-12	5622-12	8964-12	1415-13
23	1968-11	3529-11	6234-11	1086-12	1866-12	3167-12	5310-12	8799-12	1442-13	2339-13
24	2296-11	4264-11	7793-11	1403-12	2489-12	4355-12	7522-12	1283-13	2163-13	3605-13
25	2480-11	4776-11	9039-11	1683-12	3086-12	5574-12	9929-12	1745-13	3028-13	5192-13
26	2480-11	4959-11	9735-11	1877-12	3561-12	6646-12	1222-13	2215-13	3960-13	6989-13
27	2296-11	4776-11	9735-11	1947-12	3824-12	7385-12	1403-13	2625-13	4840-13	8800-13
28	1968-11	4264-11	9039-11	1877-12	3824-12	7649-12	1503-13	2907-13	5532-13	1037-14
29	1561-11	3529-11	7793-11	1683-12	3561-12	7385-12	1503-13	3007-13	5913-13	1144-14
30	1145-11	2705-11	6234-11	1403-12	3086-12	6646-12	1403-13	2907-13	5913-13	1183-14

probability of the next term in the series is .00007, which changes P to .03241 and makes the error less than .00007.

It can also be shown that the following inequalities* exist:

$$p\left(1 + \frac{ad}{(b+1)(c+1)}\right) \leq P \leq p\left(1 + \frac{ad}{(b+1)(c+1) - (a-1)(d-1)}\right),$$

where p and P are defined in (1) and (2), and ad is taken to be the smaller and bc the larger of the two products. For the example,

$$.02608\left(1 + \frac{64}{304}\right) \leq P \leq .02608\left(1 + \frac{64}{255}\right)$$

$$.03157 \leq P \leq .03263.$$

In the accompanying table, binomial coefficients, ${}_nC_r$, are given to four significant digits. Values were calculated to six significant digits, rounded off to four, and then checked against a recent source [5]. A table of logarithms of binomial coefficients, available in Hald [4] to $n = 100$, can be substituted for a table of binomial coefficients.

In the case of $r \times 2$ tables, the probability of a specified table given the column sums m and n, and row sums $(a + b)$, $(c + d)$, $(e + f)$, $\cdots$, is

$$\frac{(a+b)!}{a!\,b!}\frac{(c+d)!}{c!\,d!}\frac{(e+f)!}{e!\,f!}\cdots\frac{m!\,n!}{N!} = \frac{{}_{a+b}C_a\ {}_{c+d}C_c\ {}_{e+f}C_e}{{}_NC_m}.$$

Here also the use of binomial coefficients is economical. The procedure is still laborious, however, since it is necessary to lay out all of the possible tables to find those which are equally or less probable than the one in question [3].

REFERENCES

[1] Federighi, E. The use of chi square in small samples. Amer. *soc. Rev.*, 1950, **15**, 777-779.
[2] Fisher, R. A. Statistical methods for research workers. New York: Hafner, 1954.
[3] Freeman, G. H. and Halton, J. H. Note of an exact treatment of contingency, goodness of fit, and other problems of significance. *Biometrika*, 1951, **38**, 141-149.
[4] Hald, A. Statistical tables and formulas. New York: Wiley, 1952.
[5] Miller, J. C. P., editor. Tables of binomial coefficients. Royal Society mathematical tables Vol. 3. Cambridge: University Press, 1954.

Manuscript received 3/14/55

Revised manuscript received 11/29/55

*We are indebted to Professor John W. Tukey for calling our attention originally to a similar set of inequalities, which we modified slightly.

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

A STOCHASTIC MODEL FOR ROTE SERIAL LEARNING

RICHARD C. ATKINSON*

INDIANA UNIVERSITY†

A model for the acquisition of responses in an anticipatory rote serial learning situation is presented. The model is developed in detail for the case of a long intertrial interval and employed to fit data where the list length is varied from 8 to 18 words. Application of the model to the case of a short intertrial interval is considered; some predictions are derived and checked against experimental data.

This paper represents a preliminary attempt at quantitative theorizing in the area of rote serial learning. The model is applicable to experimental situations employing the anticipation method [6] and deals with the acquisition of correct responses, anticipatory responses, perseverative responses, and failures-to-respond. In addition, direct applicability of the model is limited to situations restricted as follows: (*a*) moderate presentation rate, (*b*) dissimilar intralist words, (*c*) familiar and easily pronounced words. The explanation for these restrictions is considered later.

Model

The model makes use of the conceptual formulation of the stimulating situation introduced by Estes [3] and elaborated by Estes and Burke [4]. The general assumptions are: (*a*) the effect of a stimulating situation upon an organism is made up of many component events; (*b*) when a situation is repeated over a series of trials, any one of these component stimulating events may occur on some trials and fail to occur on others. Rather than review the rationale of these assumptions, the reader is referred to the Estes-Burke paper which is helpful to an understanding of the present work.

Figure 1 schematically presents the rote serial learning situation. The successive word exposures in a list of $r + 1$ words are indicated by $W_1 , W_2 , \cdots , W_r , W_{r+1}$ where W_1 is the cue for S's first anticipation on each run through the list. R_i' represents a hypothesized covert response associated with the $i + 1$st word presentation; the response of "reading" W_{i+1}. On the other hand, $R_i(i)$ is the response recorded by the experimenter to the ith word presentation and can be either (*a*) a correct anticipation

*The author wishes to thank Professors C. J. Burke and W. K. Estes for advice and assistance in carrying out this research.

†Now at Stanford University.

of the $i + 1$st word when $j = i$, (*b*) an incorrect anticipation when $j \neq i$, or (*c*) a failure-to-respond when the j subscript is omitted. (Symbols and their meanings are listed in Appendix B.)

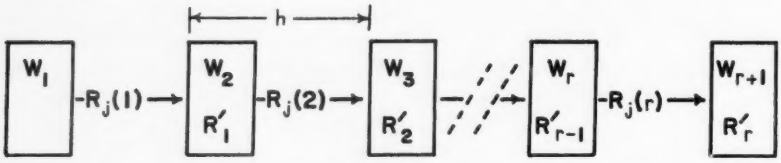

FIGURE 1

Schematic representation of the anticipatory rote serial learning situation.

A period h is defined as the time of a single word exposure, and a trial refers to one run through the list. Since the removal of one word is followed immediately by the presentation of the next, a trial is of time $h(r + 1)$. The intertrial interval is represented as a series of k subintervals each of length h; thus, the intertrial interval is of time kh. When there are $r + 1$ words in a list, the list length is designated as r; this reflects the fact that the $r + 1$st word is not a cue for an anticipatory response.

The ith word presentation is represented conceptually as a set of stimulus elements S_i where the sets are pairwise disjoint, and hence the intersection of the $r + 1$ sets is the null set. The number of elements in S_i is N, where N is invariant over i, and a parent set S^* is defined such that the union of the $r + 1$ sets is a subset of S^*. On a given presentation of the ith word a sample of elements from S_i is effective; the likelihood of any element from S_i being in the sample is θ_i where $0 \leq \theta_i \leq 1$. (Derivations presented in this paper are carried out under the simplifying assumption that all elements in S_i are equally likely to occur on any trial.) Therefore, given the ith word presentation, a sample is drawn from S_i of size $N\theta_i$.

Conditional relations, or connections, between response classes and stimulus elements are defined as in other papers on statistical learning theory. The response classes R_1 , R_2 , $\cdots$, R_r , and $\bar{R}$ (failure-to-respond) define a partition of S^* into subsets $S^*_{R_1}$, $S^*_{R_2}$, $\cdots$, $S^*_{\bar{R}}$. Elements in $S^*_{R_i}$ are said to be conditioned to the response class R_i etc. The concept of a partition implies that every element of S^* must be conditioned to either R_1 , R_2 , $\cdots$, or $\bar{R}$, but that no element may be conditioned to more than one. For each element in S_i a quantity $F(i; j; n)$ is defined which represents the probability that an element from set S_i is conditioned to response class R_j at the start of trial n. At times this notation is unnecessarily detailed; the abbreviation $C(i; n)$ is introduced to designate the probability that an element from S_i is conditioned at the start of trial n to a correct anticipatory response.

The anticipatory response at position i on trial n is assumed to be a function of the stimulus elements sampled from S_i on that trial. Specifically, the probability of $R_j(i)$ is the ratio of the number of sampled elements from S_i conditioned to the response class R_j to the number of elements sampled from S_i . Since θ_i is constant for all elements in S_i , the probability of $R_j(i)$ on trial n is the expected value of $F(i; j; n)$.

For each element sampled from S_i on trial n it is postulated that there is: (*a*) a probability λ that the element is returned to S* during the *h*-interval immediately following the one in which it was sampled; (*b*) a probability $\lambda(1 - \lambda)$ that it it is returned to S* during the second *h*-interval following the one in which it was sampled; (*c*) a probability $\lambda(1 - \lambda)^2$ that it is returned to S* during the third *h*-interval following the one in which it was sampled; and so on. The probability that an element will be eventually returned to S* is unity since

$$\sum_{x=0}^{\infty} \lambda(1 - \lambda)^x = 1. \tag{1}$$

The phrase "available at position i" is used to refer to an element sampled from some set and not yet returned to S* during the *h*-interval in which W_i is presented. The notion of an element being available at a position other than the one at which it was sampled is one way of formalizing the concept of trace stimuli. Parenthetically, note that the probability of an anticipatory response at position i is defined in terms of the stimulus elements sampled from S_i and is not affected by elements which are available at position i but sampled from a stimulus set other than S_i .

The conditioned status of elements sampled from S_i upon their return to S* depends on the anticipatory response made at position i. If a sample is drawn from S_i which elicits a correct anticipatory response, $R_i(i)$, then all elements in the sample become conditioned to the response class R_i and, independent of the time that an element is available, are returned to S* conditioned to that response class. On the other hand, if the sample elicits a response, other than a correct one, all elements in the sample revert to being conditioned to the response class $\bar{R}$, and there is a specified probability that the elements will be conditioned to the R_i' responses which occur before they are returned to S*. That is, given an incorrect anticipation or a failure-to-respond, all sampled elements become conditioned to the response class $\bar{R}$ and then: (*a*) a proportion β of the sampled elements are conditioned to the response class R_i when R_i' occurs, and $(1 - \beta)$ remain unchanged; (*b*) λ of the elements are then returned to S* and $(1 - \lambda)$ remain available during the next *h*-interval where, again, β of the remaining elements are conditioned to the response class R_{i+1} when R_{i+1}' occurs, and $(1 - \beta)$ remain as they were in the previous interval; (*c*) $\lambda(1 - \lambda)$ are now returned to S* and $(1 - \lambda)^2$ are carried on where β are connected to the response class R_{i+2}

when R'_{i+2} occurs and $(1 - \beta)$ remain as they were in the previous interval; and so on.

Finally, it is assumed that nothing which occurs during the intertrial interval will change the conditional status of the elements not yet returned to S* at the beginning of this interval. That is, elements returned during h-intervals of the intertrial interval have the same conditional status as elements returned in the last h-interval of the list presentation.

More generally stated, if a sample of elements elicits a response which is confirmed as correct (reinforced), then each element in the sample becomes conditioned to that response and will remain conditioned unless the element is sampled at some later trial, and this new sample elicits an incorrect response. If a sample leads to an incorrect response, then the elements in the sample revert to being conditioned to the response class $\bar{R}$ and have a probability β of being conditioned to the response class R_i associated with the R'_i responses which occur before the element is returned to S*. The conditioning proportion β can be interpreted as the probable occurrence of the implicit response R'_i to the $i + 1$st word presentation. This interpretation does not affect the quantitative formulation of the model.

The present analysis of serial responding requires a modification of the notion of a sampling constant introduced in other papers on statistical learning theory. θ_i is postulated to be a function of the number and order of the words that have preceded the ith word. Once again, consider intervals of time h. If the word exposure has been preceded by an infinite number of h-intervals which do not contain word exposures, then the sampling constant is θ_1 ; if, on the other hand, the word exposure has been preceded by an infinite number of h-intervals each of which contained a word exposure, the sampling constant is θ_∞ . Let $c = \theta_1 - \theta_\infty$, where $c \geq 0$ and, necessarily, $c \leq 1$. Further, designate a decay constant η such that $0 \leq \eta \leq 1$. If a series of successive word exposures occur, and are preceded by an infinite number of h-intervals which do not contain word exposures, then (a) the sampling constant associated with the second word exposure is $\theta_1 - c\eta$; (b) the sampling constant associated with the third word is $\theta_1 - c[\eta + \eta(1 - \eta)]$; ($c$) the sampling constant for the fourth word is $\theta_1 - c[\eta + \eta(1 - \eta) + \eta(1 - \eta)^2]$; and so on. Thus, if the intertrial interval is infinite (i.e., each run through the list is preceded by an infinite number of h-intervals which do not contain word exposures), the sampling constant associated with set S_i on any run through the list is

$$\theta_i = \theta_1 - c[1 - (1 - \eta)^{i-1}]. \tag{2}$$

An inspection of this equation indicates that θ, defined over list positions, has a maximum at position one and approaches $0 \leq \theta_1 - c \leq 1$ as i becomes large.

The formulation of the sampling constant requires a uniform activity

during intervals which do not contain word exposures; θ_1 is postulated to be a function of the type of activity.

The equations specified by the above assumptions can now be written. Consider the case in which the intertrial interval is "long," for purposes of the model infinite. This case proves to be simpler than that in which the intertrial interval is "short" because in the infinite interval all elements sampled from S_i on trial n are returned to S* before the beginning of trial $n + 1$ (see equation 1). (Perseverative errors are not possible for the infinite intertrial interval, and their consideration is deferred until discussion of the short interval case.)

Given a list length r and an infinite intertrial interval, the expected values of the probabilities of correct anticipatory responses on trial $n + 1$ to the exposure of W_r , W_{r-1} , and W_{r-2} are

$$C(r; n+1) = (1 - \theta_r)C(r; n) + \theta_r\{C(r; n) + [1 - C(r; n)]\beta\}, \tag{3}$$

$$\begin{aligned} C(r-1; n+1) &= (1 - \theta_{r-1})C(r-1; n) \\ &+ \theta_{r-1}\{C(r-1; n) + [1 - C(r-1; n)][\lambda\beta + (1-\lambda)\beta(1-\beta)]\}, \end{aligned} \tag{4}$$

$$\begin{aligned} C(r-2; n+1) &= (1 - \theta_{r-2})C(r-2; n) + \theta_{r-2}\{C(r-2; n) \\ &+ [1 - C(r-2; n)][\lambda\beta + \lambda(1-\lambda)\beta(1-\beta) + (1-\lambda)^2\beta(1-\beta)^2]\}. \end{aligned} \tag{5}$$

More generally,

$$C(i; n+1) = (1 - \theta_i)C(i; n) + \theta_i\{C(i; n) + [1 - C(i; n)]\beta\Delta_i\}, \tag{6}$$

where

$$\Delta_i = \lambda \frac{1 - [(1-\lambda)(1-\beta)]^{r-i}}{1 - (1-\lambda)(1-\beta)} + [(1-\lambda)(1-\beta)]^{r-i}. \tag{7}$$

Inspection of (7) indicates that Δ, defined over list positions, is bounded between zero and unity. The function assumes a minimum at position one and increases as i becomes large to a maximum value of unity at position r.

The solution of difference equation (6) is

$$C(i; n) = 1 - [1 - C(i; 0)][1 - \theta_i\beta\Delta_i]^n \tag{8}$$

(cf. [5]).

Similar sets of equations (see Appendix A) can be written for the probability of an anticipatory error and failure-to-respond. However, for simplicity, analysis is limited here to $C(i; n)$.

For the typical rote serial learning situation, assume $C(i; 0) = 0$; that is, on the first run through the list S will make no correct anticipations. The probability of an error on trial n at position i is $[1 - C(i; n)]$, and the number

of errors at position i during the first $x + 1$ trials is

$$\sum_{n=0}^{x} [1 - C(i; n)] = \frac{1 - [1 - \theta_i\beta\Delta_i]^{x+1}}{\theta_i\beta\Delta_i}. \tag{9}$$

As x becomes large this expression approaches

$$1/(\theta_i\beta\Delta_i). \tag{10}$$

Application to Data

Data have been collected for different list lengths with a one-minute intertrial interval [1]. The lists were composed of familiar and easily pronounced two-syllable adjectives; no two words possessed similar meaning or phonetic construction. The data on total number of errors over the first 16 trials at each list position are presented in Figure 2. Each curve is based

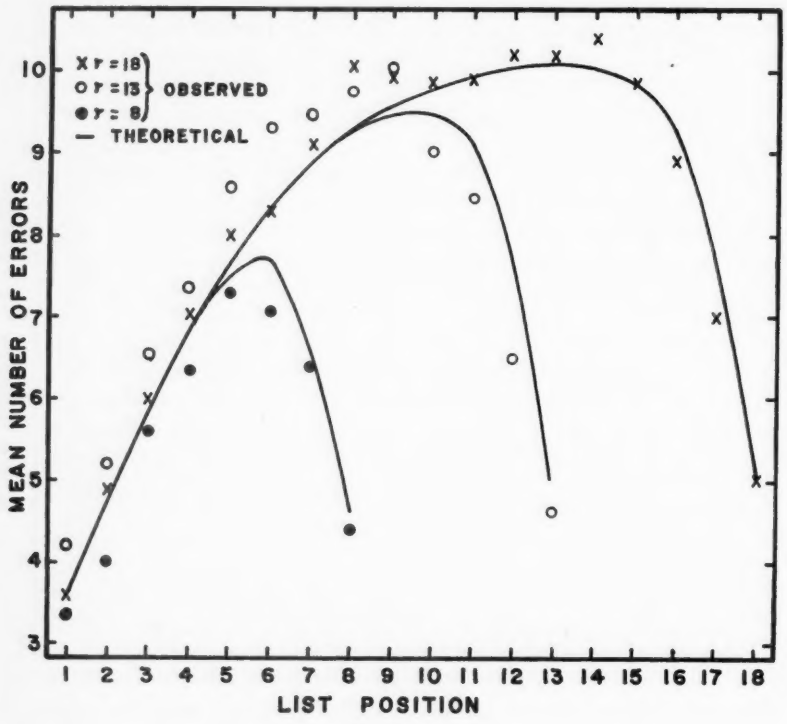

FIGURE 2

Theoretical and observed values of mean number of errors by serial positions over the first 16 trials for lists of length 8, 13, and 18.

on the records of 42 *S*s obtained in a situation employing a latin square design. Evidence on intertrial interval [1] suggests that the one-minute period experimentally approximates the theoretical infinite intertrial interval. Therefore equations (2) and (10) are applicable. These equations were employed to provide a visual fit to data for the list in which r equals 18; the obtained parameter values were $\lambda = .41$, $\beta = .55$, $\theta_1 = 1.00$, $c = .64$, and $\eta = .35$. These values were substituted in equations (2) and (10) to yield predicted curves for r equal to 8 and 13. An inspection of Figure 2 indicates close agreement between predicted and observed values.

Discussion

In the introduction the class of rote serial learning experiments to which the model is presumed to apply was delimited. The reasons for these restrictions are:

(*a*) *Moderate presentation rate.* A presentation rate that is too rapid would tend to decrease the likelihood of overt verbal responses and lead to an increase in the number of failures-to-respond. Consequently the model when applied to conditions of rapid presentation would underestimate the observed number of failures-to-respond. On the other hand, the model assumes that a single sample is drawn from S_i during the W_i exposure, an assumption which is to depend on a short exposure period. Experimentally these difficulties can be resolved by a short word exposure period followed by a blank exposure during which *S* provides an anticipation or failure-to-respond. An extension of the model to the case of a rapid rate has been examined, but the equations will not be displayed here.

(*b*) *Highly dissimilar words.* It is required in the model that the S_i sets be pairwise disjoint. This simplifying assumption is suspect for any serial learning situation, but it appears to provide an adequate approximation in this restricted situation. For the case of highly similar list words a set of elements common to each S_i would be introduced; the additional problems generated in this case are not considered here.

(*c*) *Familiar and easily pronounced words.* For the model, this restriction refers to a state such that the occurrence of the hypothesized W_i—R'_{i-1} relation is invariant over trials. For nonsense syllable learning the model would require, as an additional feature, a function describing the acquisition over trials of the W_i—R'_{i-1} connection [7].

In analyzing the model, the case where the intertrial interval is long has been considered. With a short interval the equations become more complex. Now some elements sampled on trial n remain available throughout the intertrial interval and into the next run through the list. For example, assume that an element is sampled from S_{r-1} on trial n and not returned to S^* for five h-intervals; the probability of this event is $\lambda(1 - \lambda)^4\theta_{r-1}$. When

$k = 1$, the element will be returned after the occurrence of R_i' on trial $n + 1$. Consequently, there is a probability $\beta[1 - C(r - 1; n)]$ that this element is conditioned to the response class R_1 . The element, when sampled again, increases the likelihood of an R_1 anticipatory response which, at position $r - 1$, would be classified as a perseverative error. It follows that the shorter the intertrial interval the greater the number of perseverative errors. This result has been experimentally verified [1].

Appendix A

Probability of a Failure-to-Respond and an Anticipatory Error

For the case of an infinite intertrial interval the probability of a failure-to-respond at position i on trial $n + 1$ is

$$\bar{R}(i; n + 1) = (1 - \theta_i)\bar{R}(i; n) + \theta_i[1 - C(i; n)](1 - \beta)\Delta_i \ . \tag{11}$$

The solution [5, p. 584] of this difference equation is

$$\bar{R}(i; n) = (1 - \theta_i)^n\bar{R}(i; 0) + \frac{(1 - \beta)\Delta_i}{1 - \beta\Delta_i}[(1 - \theta_i\beta\Delta_i)^n - (1 - \theta_i)^n], \tag{12}$$

where $\bar{R}(i; 0)$ is the probability of a failure-to-respond on the initial run through the list. The probability of an anticipatory error is

$$A(i; n) = 1 - C(i; n) - \bar{R}(i; n). \tag{13}$$

For the typical experimental situation, assume $C(i; 0) = 0$ and $\bar{R}(i; 0) = 1$; then (13) reduces to

$$A(i; n) = \frac{1 - \Delta_i}{1 - \beta\Delta_i}[(1 - \theta_i\beta\Delta_i)^n - (1 - \theta_i)^n]. \tag{14}$$

(12) and (14) when summed over the first x trials, as was done in (9) for incorrect responses, produce functions for failures-to-respond and anticipatory errors of the form reported by Deese and Kresse [2].

Appendix B

List of Symbols and Their Meanings

$A(i; n)$	probability of an anticipatory error at position i on trial n.
β	conditioning constant associated with an incorrect anticipation.
c	$\theta_1 - \theta_\infty$.
$C(i; n)$	probability of a correct anticipation at position i on trial n.
Δ_i	function defined over i; dependent on r, λ, and β.
η	decay constant related to the decrement in θ_i as i increases.
h	time of a single word exposure.
k	number of h-intervals in the intertrial interval.

λ	probability that an available element will be returned to S* during the next h-interval.
n	number of trial.
r	list length.
R_i'	hypothesized covert response; reading W_{i+1}.
R_i	response class; overt anticipation of W_{i+1}.
$\bar{R}$	response class; failure-to-respond.
$R_j(i)$	R_j recorded by experimenter to W_i.
$\bar{R}(i; n)$	probability of a failure-to-respond at position i on trial n.
S*	set of stimulus elements of which all S_i are subsets.
S_i	set of stimulus elements associated with W_i.
θ_i	probability of sampling an element from S_i when W_i occurs.
W_i	ith word presentation, where W_1 is cue for first anticipation.

REFERENCES

[1] Atkinson, R. C. An analysis of rote serial position effects in terms of a statistical model. Unpublished doctor's dissertation, Indiana Univ., 1954.

[2] Deese, J. and Kresse, F. H. An experimental analysis of the errors in rote serial learning. *J. exp. Psychol.*, 1952, **44**, 199-202.

[3] Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94-107.

[4] Estes, W. K. and Burke, C. J. A theory of stimulus variability. *Psychol. Rev.*, 1953, **60**, 276-286.

[5] Jordan, C. Calculus of finite differences. New York: Chelsea, 1950.

[6] McGeoch, J. A. and Irion, A. L. The psychology of human learning. New York: Longmans, Green, 1952.

[7] Noble, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333-337.

Manuscript received 1/12/56

Revised manuscript received 2/22/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

A COMPUTATIONAL PROCEDURE FOR TAU CORRELATION*

DESMOND S. CARTWRIGHT

UNIVERSITY OF CHICAGO

The *tau* coefficient is defined, and a computational procedure for tied ranks is described. The procedure maintains continuous computational checks, saves labor, and particularly facilitates the use of *tau* with large samples. It is also shown how *tau* correlation may be applied to *Q*-sorts with any shape of forced distribution or with unforced distributions.

It frequently happens in psychological research that the only appropriate test for an hypothesis is one that does not assume normality for the variates concerned. Where correlation is at issue, a non-paramatric coefficient is required. Spearman's *rho* is well known. Kendall's *tau* [3], a newer method of rank correlation, has several advantages over *rho*. Most important is the fact that the significance of a sample *tau* can be accurately evaluated on the basis of the normal probability integral for $n > 10$, while for $n \leq 10$ exact tables are available. The chief disadvantage of *tau* is the computational labor involved. This paper presents a method of computation designed especially for large samples and multiple ties.

Definition

Among n individuals there are $n(n-1)/2$ relations between pairs. If the individuals be ranked on two variates, each pair can agree or differ as to the order of ranks. *Tau* is defined as

$$\tau = \frac{P - Q}{n(n-1)/2}, \tag{1}$$

where n is the number of individuals, P is the number of pairs having the same rank order on both variates, and Q is the number of pairs having inverse orders. If the two rankings agree perfectly, then $P = n(n-1)/2$, $Q = 0$, and $\tau = +1.00$. If one ranking is a perfect inversion of the other, then $Q = n(n-1)/2$, $P = 0$, and $\tau = -1.00$.

*The procedure described was developed in connection with research at the Counseling Center, University of Chicago. The research is supported by a grant (PHS M 903) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

Alternate Formulas

Kendall has given formulas equivalent to (1) which reduce computation labor. For example,

$$\tau = 1 - \frac{2Q}{n(n-1)/2}. \tag{2}$$

The transformation from (1) to (2) is

$$\tau = \frac{P-Q}{n(n-1)/2} = \frac{P+Q-2Q}{n(n-1)/2} = 1 - \frac{2Q}{n(n-1)/2}, \tag{3}$$

which holds when $P + Q = n(n-1)/2$. This is true if, and only if, there are no ties. Tied pairs can generate neither an agreement nor an inversion. Tied pairs can generate only zero scores. Let the sum of these be Z. In the tied case $n(n-1)/2 = P + Q + Z$. Hence the transformation (3) is impossible. It is for this reason that Bright's recent procedure for computing τ [1] is inappropriate for the tied case. Only formula (1) is appropriate here.

Procedure

If only one ranking contains ties, arrange the paired ranks as nearly as possible in natural order from left to right on the tied variate. Bracket each tied set. An example follows, with R_a the tied variate:

$$\begin{array}{lcccccccc}
R_a: & 1 & [3 & 3 & 3] & 5 & [7 & 7 & 7] \\
R_b: & 3 & [4 & 2 & 7] & 6 & [1 & 5 & 8]
\end{array}$$

$$P = 5 + 3 + 3 + 1 + 1 + 0 + 0 + 0 = 13$$

$$Q = 2 + 1 + 1 + 3 + 2 + 0 + 0 + 0 = 9$$

$$Z = 0 + 2 + 1 + 0 + 0 + 2 + 1 + 0 = 6$$

$$\text{Check } \sum = P + Q + Z = 7 + 6 + 5 + 4 + 3 + 2 + 1 + 0 = 28$$

Certain rules of procedure may be set out as follows:

(1) There is an *agreement*, at any given number in R_b, for every *larger* number to the right of it which is *not* in the same bracket.

(2) There is an *inversion*, at any given number in R_b, for every *smaller* number to the right of it which is *not* in the same bracket.

(3) There is a *zero*, at any given bracketed number in R_b, for every number to the right of it *in the same bracket.*

Every given pair of individuals generates either an agreement (larger number to the right), an inversion (smaller number to the right), or a zero

(in the same bracket). For every ith individual, proceeding from left to right, there will be $n - i$ to the right, each of which must generate an agreement or an inversion or a zero. A check row is affixed, with entries $n - i$; for each individual, (agreements + inversions + zeros) = $n - i$. Also for the totals, $P + Q + Z = n(n - 1)/2 = \sum(n - i)$.

If there are ties in both rankings, arrange the paired ranks as nearly as possible in natural order from left to right on one tied variate R_a. Bracket each set tied on R_a. Add one more rule to the three given above:

(4) There is a zero, at any given number in R_b, for *every* equal number to the right of it which is *not* in the same bracket.

R_a :	1	2	[5	5	5	5	5]	8	[10	10	10]	12	13
R_b :	4	5	[2	6	8	8	8]	13	[11	2	2]	11	11
Z_a =	0	0	4	3	2	1	0	0	2	1	0	0	0
Z_b =	0	0	2	0	0	0	0	0	2	0	0	1	0

P= 9+ 8+ 4+4+4+4+4+0+0+2+2+0+0=41

Q= 3+ 3+ 0+2+2+2+2+5+0+0+0+0+0=19

$Z=Z_a+Z_b$= 0+ 0+ 6+3+2+1+0+0+4+1+0+1+0=18

Check $\sum=P+Q+Z=12+11+10+9+8+7+6+5+4+3+2+1+0=78$

For clarity the zero scores arising from rules (3) and (4) are separated in tabulation. Zeros arising from (3) are entered in the row Z_a. Zeros arising from (4) are entered in row Z_b. The row Z is given by $Z_a + Z_b$.

Our second example has $n = 13$. With 78 relations between pairs to be examined, computation is already laborious. A short-cut is provided by the following method.

After setting out R_a and R_b as before, R_b is given a separate tabular form. This is called a "B-chart". The stub-head contains all ranks and mid-ranks of R_b written in natural order. For any set of ties only one mid-rank is represented. The top row of the table proper shows the number of individuals in each set. Let that number be u. Columns are then filled in with entries decreasing successively by 1 from u until unity is reached. The complete B-chart for our second example appears like this:

Rank or mid-rank :	2	4	5	6	8	11	13
Number of individuals:	3	1	1	1	3	3	1
	2				2	2	
	1				1	1	

The B-chart gives an orderly representation of R_b . The sum of numbers in the top row is equal to n. Each entry in the top row gives the number of individuals having the rank or mid-rank shown in the stub-head. Thus there are three individuals with mid-rank of 8. The chart shows at a glance how many individuals have a rank number *larger* than a given number. Thus the sum of numbers in the top row to the *right* of column 8 is equal to 4. It shows how many individuals have a rank number *smaller* than a given number. Thus the sum of numbers in the top row to the *left* of column 8 is equal to 6.

Computation by the rules listed above proceeds from left to right. Before applying the computation rules to any given number, we strike out its appropriate entry in the B-chart. Thus, for any given number in R_b , the chart will show how many numbers *to the right of that number in* R_b are larger than, smaller than, or equal to the given number. Thus rapid computation for rules (1), (2) and (4) is possible. Rules for using the B-chart follow:

(5) For any given *non-bracketed* number in R_b , enter the B-chart under the column-head with that number, and strike out the topmost visible figure in the column. The sum of topmost visible figures (i.e., not struck out) in all columns to the right of that entered gives the number of agreements for rule (1). The sum of topmost visible figures in all columns to the left of that entered gives the number of inversions for rule (2). In the column entered, the topmost visible figure (after the strike-out) gives the number of zeros for rule (4).

(6) Before computing by rules (1), (2) and (4) for any given *bracketed* number in R_b , enter the B-chart under the column-head with that number and strike out the topmost visible figure in the column. Repeat for each number in the bracket until all appropriate strike-outs are made. Computations then proceed as in rule (5).

Use of rules (5) and (6) will be illustrated with the second example. The first number in R_b is 4. By rule (5) the entry 1 in column 4 of the B-chart is struck out. Topmost visible figures in columns to the right of 4 are summed, giving 9 agreements by rule (1). Summing to the left gives 3 inversions by rule (2). There are no zeros by rule (4). The second number in R_b is 5. By rule (5) the entry 1 in column 5 is struck out. Summing to the right gives 8 agreements by rule (1). Summing to the left gives 3 inversions by rule (2). There are no zeros by rule (4). The third number in R_b is 2, and it is bracketed. By rule (6) the entry 3 in column 2 of the B-chart is struck out. Other numbers in the same bracket are 6, 8, 8 and 8. The entry 1 in column 6, and the entries 3, 2 and 1 in column 8 are all struck out. Just before computing for the third number in R_b , the B-chart would look like this:

Rank or mid-rank :	2	4	5	6	8	11	13
Number of individuals:	–	–	–	–	–	3	1
	2				–	2	
	1				–	1	

For the third number in R_b, which is 2, summing to the right gives 4 agreements by rule (1); there are no inversions by rule (2); there are two zeros by rule (4), as given by the topmost visible figure in column 2 itself.

Rule (3) does not require use of the B-chart. In a bracketed set of t numbers in R_b, the ith number generates $t - i$ zeros by rule (3). This number of zeros is entered directly in row Z_a.

Formulas for Ties

The method described above yields P and Q, components of the numerator for τ. Kendall gives two alternative denominators for use with ties. The first, τ_a, implies an untied criterion ranking. Ties indicate departures from this criterion even as inversions do. Hence the denominator remains $n(n-1)/2$. The second, τ_b, does not imply a criterion. The presence of ties simply reduces the maximum possible number of agreements. The denominator is accordingly reduced and made equal to the geometric mean of $P_{\max}$ for R_a and $P_{\max}$ for R_b. For any set of n untied ranks, $P_{\max} = n(n-1)/2$, and $Q_{\min} = 0$. If there are ties $Q_{\min}$ is still 0, but every pair of numbers within a tie generates a zero, and thereby reduces $P_{\max}$. Within any tie, the number of such zeros is $t(t-1)/2$. Hence for R_a, $P_{\max} = n(n-1)/2 - \sum_t [t(t-1)/2]$, where for each tie, t is the number of ranks tied, and $\sum_t$ means summation over all *sets* of ties in R_a. If we label the ties in R_b, u, then $P_{\max}$ for $R_b = n(n-1)/2 - \sum_u [u(u-1)/2]$. Then,

$$\tau_b = \frac{P - Q}{\sqrt{[n(n-1)/2] - T}\sqrt{[n(n-1)/2] - U}}, \tag{4}$$

where

$$T = \sum_t t(t-1)/2,$$

$$U = \sum_u u(u-1)/2.$$

For our second example, $T = (5 \cdot 4/2) + (3 \cdot 2/2) = 13$; $U = (3 \cdot 2/2) + (3 \cdot 2/2) + (3 \cdot 2/2) = 9$; and

$$\tau_b = \frac{41 - 19}{\sqrt{(78 - 13)}\sqrt{(78 - 9)}} = .329.$$

Application of τ to Large Samples

The procedure described in this paper may be used with large samples of any kind of appropriate data. However, illustrations will here be confined to investigations using subjective metrics.

Butler and Fiske [2] have recently argued for increased use of subjective metrics in personality assessment. They point out [2, p. 332] that a card-sort, as used in Stephenson's Q-technique [4], can be regarded as a ranking with a fixed number of ties. They assert that τ correlation is an appropriate method of analysis for such sorts.

When cards are forced into a normal distribution by the experimental instructions it seems immaterial whether the product moment or the τ correlation procedure is employed. In some investigations however, it is desired to use a forced rectangular distribution, or some other forced shape of distribution which may or may not satisfy the requirements for using a product moment coefficient. In other investigations maximum reduction of experimental constraint may be required, so that the subject is left free to distribute the cards in any way he pleases. Under such conditions the τ coefficient provides an appropriate test for hypotheses concerned with correlation.

In Table 1 a B-chart is presented for the case of forced-normal sorts of 64 cards. It will be noted that the column heads are pile numbers instead of ranks. The subject is required to sort the cards into 7 piles with the given distribution. The metric provided by the instruction may be "from most to least significant for you." The basic order relations between cards are therefore given by the pile numbers, and it is unnecessary to transform these to ranks.

Suppose R_a and R_b on the 64 cards are set out as for our second example. Suppose the first number in R_b is 4. Then, following rule (5), there are $15 + 6 + 1 = 22$ agreements by rule (1); there are $15 + 6 + 1 = 22$ inversions by rule (2); and there are 19 zeros by rule (4). Instead of taking $n - 1 = 63$ observations of the relations between the first member and all other members, only three readings are made from the chart in Table 1.

Where work is being done with forced-sorts, one blank chart of the kind shown in Table 1 can be mimeographed for all correlations on sorts having the given distribution. When the investigation is concerned with distribution-free sortings, individual charts must be prepared. For extensive work of this kind, however, it is possible to set up a generalized chart with rows $r =$ (the maximum likely number of cards sorted into any one pile), and columns $c =$ (the maximum likely number of piles). Every column is then filled with numbers descending from r successively by 1. For a particular correlation the B-chart is then drawn in ink. This procedure is exemplified in Table 2.

The distribution outlined in Table 2 gives the B-chart for a hypothetical sorting of 26 colors in terms of preference. Using this chart, and machine

TABLE 1

B-chart for a Forced-Normal Sort of 61 Items

Pile Numbers						
1	2	3	4	5	6	7
1	6	15	20	15	6	1
	5	14	19	14	5	
	4	13	18	13	4	
	3	12	17	12	3	
	2	11	16	11	2	
	1	10	15	10	1	
		9	14	9		
		8	13	8		
		7	12	7		
		6	11	6		
		5	10	5		
		4	9	4		
		3	8	3		
		2	7	2		
		1	6	1		
			5			
			4			
			3			
			2			
			1			

TABLE 2

General Chart for 8 Rankings with Unforced Sorts of 26 Items
Showing a Specific B-chart for a Particular Sort

Pile Numbers											
1	2	3	4	5	6	7	8	9	10	11	12
12	12	12	12	12	12	12	12	12	12	12	12
11	11	11	11	11	11	11	11	11	11	11	11
10	10	10	10	10	10	10	10	10	10	10	10
9	9	9	9	9	9	9	9	9	9	9	9
8	8	8	8	8	8	8	8	8	8	8	8
7	7	7	7	7	7	7	7	7	7	7	7
6	6	6	6	6	6	6	6	6	6	6	6
5	5	5	5	5	5	5	5	5	5	5	5
4	4	4	4	4	4	4	4	4	4	4	4
3	3	3	3	3	3	3	3	3	3	3	3
2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	1	1	1	1

TABLE 3

Example of Computational Routine for Tau on an Unforced Sort of 26 Items*

A	1	2	2	2	3	3	3	3	3	4	4	4	5	5	5	5	5	5	6	6	6	6	7	7	7	7
B	3	3	1	1	2	8	9	4	4	2	3	5	5	5	4	4	5	7	6	6	7	8	9	7	8	8
Z_a	0	2	1	0	4	3	2	1	0	2	1	0	5	4	3	2	1	0	3	2	1	0	3	2	1	0
Z_b	2	1	0	0	1	3	1	2	2	0	0	3	0	0	0	0	0	2	0	0	1	2	0	0	0	0
P	19	+ 19	+ 22	+ 22	+ 16	+ 1	+ 0	+ 13	+ 13	+ 14	+ 14	+ 9	+ 8	+ 8	+ 8	+ 8	+ 8	+ 4	+ 4	+ 4	+ 3	+ 1	+ 0	+ 0	+ 0	+ 0
Q	4	+ 2	+ 0	+ 0	+ 0	+ 13	+ 16	+ 2	+ 2	+ 0	+ 0	+ 2	+ 0	+ 0	+ 0	+ 0	+ 0	+ 2	+ 0	+ 0	+ 0	+ 1	+ 0	+ 0	+ 0	+ 0
Z	2	+ 3	+ 1	+ 0	+ 5	+ 6	+ 3	+ 3	+ 2	+ 2	+ 1	+ 3	+ 5	+ 4	+ 3	+ 2	+ 1	+ 2	+ 1	+ 2	+ 2	+ 2	+ 3	+ 2	+ 1	+ 0
Σ	25	+ 24	+ 23	+ 22	+ 21	+ 20	+ 19	+ 18	+ 17	+ 16	+ 15	+ 14	+ 13	+ 12	+ 11	+ 10	+ 9	+ 8	+ 7	+ 6	+ 5	+ 4	+ 3	+ 2	+ 1	+ 0

P = 218
Q = 44
Z = 63
Σ = 325

$T = \sum_t t(t-1) / 2 = 43$

$U = \sum_u u(u-1) / 2 = 28$

$$\tau_b = \frac{P - Q}{\sqrt{n(n-1)/2 - T}\ \sqrt{n(n-1)/2 - U}} = \frac{174}{287.4} = .605$$

* The B-chart for this example is shown in Table 2.

summation across the rows of P, Q, and Z, the computation of τ for two unforced sortings of 26 items can be done in less than 10 minutes. The entire routine is illustrated in Table 3.

REFERENCES

[1] Bright, H. F. A method for computing the Kendall Tau Coefficient. *Educ. psychol. Measmt.*, 1954, 14, 700-704.
[2] Butler, J. M. and Fiske, D. W. Theory and techniques of assessment. In C. P. Stone (Ed.), Annual Review of Psychology. Stanford: Annual Reviews Inc., 1955, Pp. 327-356.
[3] Kendall, M. G. Rank correlation methods. London: Griffin, 1948.
[4] Stephenson, W. The study of behavior: Q-technique and its methodology. Chicago: Univ. Press, 1953.

Manuscript received 6/20/55
Revised manuscript received 1/23/56

PSYCHOMETRIKA—VOL. 22, NO. 1
MARCH, 1957

ITERATIVE INVERSE FACTOR ANALYSIS—A RAPID METHOD FOR CLUSTERING PERSONS

BERNARD M. BASS

LOUISIANA STATE UNIVERSITY

By interchanging persons and items, iterative inverse factor analysis provides a relatively inexpensive way of clustering persons according to their patterns of response to the items. In addition to permitting the clustering of large numbers of persons, the technique enables one to determine the bases for such clustering. The items of behavior used can be heterogeneous in content and form.

Wherry and his associates have described iterative procedures for factor-analyzing large numbers of test items [1, 2, 4]. They suggest that the iterative approach would yield the same factor structure as the traditional multiple-centroid procedure. Moreover, the iterative approach would provide the factor loadings of each item rather than merely the loadings of each total test score.

The original development involves the following procedures for a test of dichotomously scored items:

1. Obtain a total score X_1 on all items for each subject.
2. Obtain the tetrachoric correlation coefficient between each item and the total score.
3. Select those items to form pool 1 which correlate highest with the score X_1.
4. Obtain a total score X_2 using all items less those items in pool 1.
5. Repeat steps 2 and 3 to obtain pool 2.
6. Iterate until all communality among items has been accounted for by the pool scores.
7. The pool scores lack independence; but the obtained oblique factor matrix can be rotated to simple structure, and, if desired, to an orthogonal solution.

The present article proposes that all the advantages of the iterative technique can be applied to *inverse* factor analysis by interchanging subjects and items. Iterative inverse factor analysis provides a means for clustering persons according to their response patterns. The behavior assessed can be measured in a variety of ways, and both sample size and the scope of behavior studied can be increased greatly with relatively little increase in analysis time.

Additional advantages may include the following: (1) The "factorial composition" of each individual person can be obtained—limited, of course, to the variety of behaviors assessed. (2) Persons can be clustered according to the pattern of their specific dichotomous responses without any initial assumptions about item relations. (3) Large numbers of persons and items can be studied. (4) The items can be mixtures of any kind of attributes and behaviors. Dichotomously grouped and quantitative data can be used simultaneously. (5) Specific items of behavior, the source or basis of "person clusters," can be determined.

Inverse iterative factor analysis proceeds as follows: For N dichotomously scored items (i.e., accept—reject):

1. Obtain the frequency with which all subjects responded accept to each specific item.

2. Order all items according to the frequency with which the accept response was given. Divide this item distribution into two equal parts, an upper and lower half.

3. Key each item according to whether it is in the upper or lower half of the distribution of acceptance.

4. Determine the frequency (X_u) with which a given subject responded accept to items in the upper half of the distribution.

5. Determine the frequency (X) with which the same subject responded accept to all N items.

6. For the given subject, enter X_u and X in Table 1 as shown, where:

   (a) X_u is the number of times the given subject responded accept to the half of the items to which all subjects responded accept most frequently.
   (b) X is the frequency with which the given subject responded accept to all items.
   (c) N is the total number of items, constant for all subjects.

7. Obtain the totals $N/2$ for each row of Table 1, by dividing in half the total number of items N. Obtain the number of items to which the given subject responded reject $N - X$ by subtraction. Complete the remaining cells of the four-fold table by subtraction.

8. Obtain the tetrachoric correlation between the given subject's and all subjects' tendencies to respond accept to the same N items, using the data of Table 1.

9. Order all subjects according to their respective tetrachoric correlations obtained in step 8. Select those with the highest correlations for pool 1 using an arbitrary cut-off value, for example, the lowest correlation statistically significant from zero at the 1 per cent level.

10. Repeat steps 1 through 9 after eliminating subjects in pool 1 from

TABLE 1

Table for Computing Tetrachoric Correlation Between a Given Subject and All Subjects

	A Given Subject's Responses		
	"Accept"	"Reject"	Both
Upper 50% of Items "Accepted" by All Subjects	X_u	$N/2 - X_u$	N/2
Lower 50% of Items "Rejected" by All Subjects	$X - X_u$	$N/2 + X_u - X$	N/2
All Items	X	N - X	N

the sample. X remains constant for a given subject. Note that N remains constant for all subjects during the iterations, while X_u varies with each successive iteration for each individual. None of the items nor any of the responses are ever eliminated from consideration as the iteration proceeds.

11. Continue iterations until the correlations between each person and each pool of persons have been obtained. This is the unrotated oblique factor matrix.

12. Rotate to simple structure.

Adding subjects to the analysis merely serves to augment the work load arithmetically, whereas a geometric increase would be involved if traditional approaches were used. Traditionally, the clustering of, say 100 persons would require approximately four times the work of clustering 50 persons. Clustering 1,000 persons would be unmanageable for most experiments, since the work would become 400 times as great as for clustering 50 persons. Moreover, assuming that IBM cards or scoring sheets are employed in the procedures outlined here, adding items of behavior to be scored entails relatively little work.

REFERENCES

[1] Wherry, R. J. and Gaylord, R. H. The concept of test and item reliability in relation to factor pattern. *Psychometrika*, 1943, **8**, 247-269.

[2] Wherry, R. J., Perloff, R., and Campbell, J. T. An empirical verification of the Wherry-Gaylord iterative factor analysis procedure. *Psychometrika*, 1951, **16**, 67-74.

[3] Wherry, R. J. and Winer, B. J. A method for factoring large numbers of items. *Psychometrika*, 1951, **18**, 161-179.

Manuscript received 7/25/55

Revised manuscript received 1/16/56